

**Selección del diseño muestral en investigaciones sociales**

**Aplicación a beneficiarios de créditos estatales en la ciudad de Bogotá D.C.**

Adriana Elizabeth Contreras Roldan  
Carlos Andrés Mesías Ocaña  
Hember Manuel Augusto Rodríguez Rodríguez  
Bogotá D.C., 2017

Fundación Universitaria Los Libertadores  
Facultad de Ciencias Básicas  
Especialización Estadística Aplicada

## **Selección del diseño muestral en investigaciones sociales**

### **Aplicación a beneficiarios de créditos estatales en la ciudad de Bogotá D.C.**

Presentado por:

Adriana Elizabeth Contreras Roldan  
Carlos Andrés Mesías Ocaña  
Hember Manuel Augusto Rodríguez Rodríguez  
Bogotá D.C., 2017

Director:

SEBASTIEN LOZANO FORERO

Trabajo de Grado para optar el título de  
Especialista en Estadística Aplicada

Fundación Universitaria Los Libertadores  
Facultad de Ciencias Básicas  
Especialización Estadística Aplicada

Nota de aceptación

---

---

---

---

---

---

Presidente del Jurado

---

Jurado

---

Jurado

---

Bogotá D.C \_\_\_\_\_ de \_\_\_\_\_

## Tabla de Contenido

Resumen .....	10
Summary.....	11
Introducción.....	12
Objeto de estudio.....	13
Presentación del problema .....	13
Justificación del problema.....	14
Objetivo de la investigación.....	15
Objetivo general. ....	15
Objetivos específicos.....	15
Marco teórico.....	16
Antecedentes de la investigación .....	16
Análisis de variables categóricas.....	16
Tablas de contingencia. ....	18
Estimador Chi-cuadrado de Pearson. ....	19
Diseños muestrales.....	20
Muestreo Aleatorio Simple. ....	21
Muestreo Probabilístico Estratificado. ....	21
Muestreo Estratificado Proporcional.....	21
Muestreo por Conglomerados. ....	22
Muestreo en etapas.....	22
Muestreo Complejo.....	22
Eficiencia del Diseño Muestral .....	22
Metodología de la investigación.....	25
Tipo de estudio.....	25
Población objeto.....	25
Unidades de análisis.....	25
Unidad de muestreo.....	25
Marco muestral.....	25
Variables .....	25
Análisis descriptivo de las variables .....	28
Relaciones de dependencia.....	28
Selección del diseño muestral .....	28
Cálculo del tamaño de muestra .....	29
Selección de la muestra.....	29

Eficiencia del diseño muestral.....	29
Resultados de la investigación.....	30
Análisis descriptivo.....	30
Tipo de emprendimiento o unidad productiva. ....	30
Sector económico. ....	30
Organización jurídica. ....	31
Número de trabajadores. ....	31
Localidad/Municipio. ....	31
Rango de edad. ....	32
Sexo.....	33
Nivel educativo. ....	33
Estrato socioeconómico.....	34
Monto desembolsado.....	34
Relaciones de dependencia con la variable MONTO DESEMBOLSADO .....	35
Monto desembolsado categorizado por sector económico.....	35
Monto desembolsado categorizado por número de trabajadores. ....	36
Monto desembolsado categorizado por zona geográfica.....	36
Monto desembolsado categorizado por rango de edad.....	37
Monto desembolsado categorizado por sexo.....	37
Monto desembolsado categorizado por estrato socioeconómico. ....	38
Variables explicativas .....	38
Sexo vs estrato socioeconómico.....	38
Sexo vs Sector económico.....	39
Sexo vs Número de trabajadores. ....	39
Tamaño de muestra .....	40
Tamaño de muestra para M.A.S.....	40
Tamaño de muestra para muestreo estratificado. ....	40
<i>Estratificación por sexo.</i> ....	40
<i>Estratificación por sector económico.</i> .....	41
Tamaño de muestra para muestreo bietápico por conglomerados y estratificado. ....	42
Planes muestrales y estadísticos.....	43
Muestreo aleatorio simple (M.A.S.).....	43
Muestreo estratificado. ....	45
Muestreo estratificado por sexo (M.E.S.).....	45
<i>Muestreo estratificado por sector económico (M.E.S.E.).</i> ....	48
<i>Muestreo complejo por conglomerados y estratificado.</i> ....	50
<i>Muestra 01: Conglomerada por zona geográfica y estratificada por sector económico.</i> ....	50

<i>Muestra 02: conglomerada por zona geográfica y estratificada por sexo.</i> .....	52
Discusión de diseños muestrales .....	55
Conclusiones de la investigación.....	57
Referencias .....	58

## Lista de tablas

TABLA 1 ESTADÍSTICA DESCRIPTIVA E INFERENCIAL. ....	17
TABLA 2 TIPOS DE VARIABLES. ....	18
TABLA 3 TABLA DE CONTINGENCIA PARA LAS VARIABLES SEXO Y ESTRATO SOCIOECONÓMICO. ....	19
TABLA 4 DESCRIPCIÓN DE LAS VARIABLES, PARTE 1. ....	26
TABLA 5 DESCRIPCIÓN DE LAS VARIABLES, PARTE 2. ....	27
TABLA 6 TABLA DE FRECUENCIAS PARA LA VARIABLE TEUP. ....	30
TABLA 7 TABLA DE FRECUENCIA PARA LA VARIABLE SEEC- ....	30
TABLA 8 TABLA DE FRECUENCIAS PARA LA VARIABLE NUTR (AGRUPADA). ....	31
TABLA 9 TABLA DE FRECUENCIAS PARA LA VARIABLE ZONA GEOGRAFICA. ....	32
TABLA 10 TABLA DE FRECUENCIAS PARA VARIABLE EDAD. ....	32
TABLA 11 TABLA DE FRECUENCIAS PARA VARIABLE SEXO. ....	33
TABLA 12 TABLA DE FRECUENCIAS PARA VARIABLE NIVEL EDUCATIVO. ....	33
TABLA 13 TABLA DE FRECUENCIAS PARA VARIABLE ESTRATO SOCIOECONOMICO. ....	34
TABLA 14 MEDIDAS DE TENDENCIA Y VARIABILIDAD PARA LA VARIABLE MONTO DESEMBOLSADO. ....	34
TABLA 15 TABLA DE FRECUENCIAS PARA VARIABLE MONTO DESEMBOLSADO. ....	35
TABLA 16 PRUEBAS DE INDEPENDENCIA PARA LAS VARIABLES SEEC Y MODE. ....	35
TABLA 17 PRUEBAS DE INDEPENDENCIA PARA LAS VARIABLES NUTR Y MODE. ....	36
TABLA 18 PRUEBAS DE INDEPENDENCIA PARA LAS VARIABLES ZONA GEOGRAFICA Y MODE. ....	36
TABLA 19 PRUEBAS DE INDEPENDENCIA PARA LAS VARIABLES EDAD Y MODE. ....	37
TABLA 20 PRUEBAS DE INDEPENDENCIA PARA LAS VARIABLES SEXO Y MODE. ....	37
TABLA 21 PRUEBAS DE INDEPENDENCIA PARA LAS VARIABLES ESEC Y MODE. ....	38
TABLA 22 PRUEBAS DE INDEPENDENCIA PARA LAS VARIABLES ESEC Y SEXO. ....	38
TABLA 23 PRUEBAS DE INDEPENDENCIA PARA LAS VARIABLES SEEC Y SEXO. ....	39
TABLA 24 PRUEBAS DE INDEPENDENCIA PARA LAS VARIABLES SEXO Y NUTR. ....	39
TABLA 25 TAMAÑO DE MUESTRA PARA MUESTREO ESTRATIFICADO POR SEXO. ....	41
TABLA 26 TAMAÑO DE MUESTRA PARA MUESTREO ESTRATIFICADO POR SECTOR ECONOMICO. ....	41
TABLA 27 TAMAÑO DE MUESTRA PARA SEGUNDA ETAPA DE MUESTREO BIETÁPICO ESTRATIFICADO POR SEXO. ....	42
TABLA 28 TAMAÑO DE MUESTRA PARA SEGUNDA ETAPA DE MUESTREO BIETÁPICO ESTRATIFICADO POR SECTOR ECONOMICO. ....	42
TABLA 29 RESUMEN PARA MUESTREO ALEATORIO M.A.S. ....	43
TABLA 30 MEDIDAS DEL ERROR DE MUESTREO PARA M.A.S. POR CATEGORÍAS DE MONTO DESEMBOLSADO. ....	43
TABLA 31 RESUMEN PARA MUESTREO ESTRATIFICADO POR SEXO (M.E.S.). ....	46
TABLA 32 MEDIDAS DEL ERROR DE MUESTREO PARA M.E.S. CATEGORIZADO POR MONTO DESEMBOLSADO. ....	46
TABLA 33 RESUMEN PARA MUESTREO ESTRATIFICADO POR SECTOR ECONOMICO. ....	48
TABLA 34 MEDIDAS DEL ERROR DE MUESTREO PARA M.E.S.E. CATEGORIZADO POR MONTO DESEMBOLSADO. ....	48
TABLA 35 RESUMEN PARA LA ETAPA 1 CONGLOMERADOS POR ZONA GEOGRAFICA. ....	50
TABLA 36 RESUMEN PARA LA ETAPA 2 ESTRATOS POR SECTOR ECONOMICO. ....	50
TABLA 37 MEDIDAS DE ERROR PARA MUESTREO BIETÁPICO ESTRATIFICADO POR SECTOR	

ECONOMICO.....	51
TABLA 38 LA ETAPA 1 CONGLOMERADOS POR ZONA GEOGRAFICA. ....	52
TABLA 39 RESUMEN PARA LA ETAPA 2 ESTRATIFICADA POR SEXO. ....	52
TABLA 40 MEDIDAS DE ERROR PARA MUESTREO BIETÁPICO ESTRATIFICADO POR SEXO. ....	53



## Lista de figuras

FIGURA 1 CLASIFICACIÓN DE LAS VARIABLES ESTADÍSTICAS. ....	17
FIGURA 2 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DEL M.A.S., VARIABLES MODE Y NUTR. ....	44
FIGURA 3 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DEL M.A.S., VARIABLES MODE Y SEEC. ....	44
FIGURA 4 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DEL M.A.S., VARIABLES MODE Y SEXO. ....	45
FIGURA 5 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DEL M.E.S., VARIABLES MODE Y SEEC. ....	47
FIGURA 6 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DEL M.E.S., VARIABLES NUTR Y SEEC. ....	47
FIGURA 7 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DEL M.E.S.E., VARIABLES MODE Y SEXO. ....	49
FIGURA 8 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DEL M.E.S.E., VARIABLES NUTR Y SEXO. ....	49
FIGURA 9 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DEL MUESTREO BIETÁPICO ESTRATIFICADO POR SECTOR ECONOMICO. ....	52
FIGURA 10 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DEL MUESTREO BIETÁPICO ESTRATIFICADO POR SEXO. ....	54
FIGURA 11 GRÁFICO DE BARRAS PARA COEFICIENTES DE VARIACIÓN DE LOS DISEÑOS MUESTRALES. FUENTE LOS AUTORES. ....	56

## Resumen

Dado que la satisfacción y la percepción son los indicadores más importante para medir la calidad de los servicios, se hace necesario aplicar, después de dos años de implementado el programa distrital de financiamiento a unidades productivas de propiedad de población vulnerable y especial principalmente, una encuesta de satisfacción y percepción con el objetivo de conocer la evaluación y opiniones que los ciudadanos beneficiados tienen del programa. Gracias a dicho planteamiento de los entes gubernamental de Colombia, el presente trabajo de investigación busca realizar la comparación de diferentes diseños muestrales para determinar aquel que encuentra la muestra más representativa de la base de datos de los beneficiarios con créditos para el fortalecimiento de sus unidades productivas en la ciudad de Bogotá.

**Palabras Clave:** Poblaciones finitas, Variables categóricas, Diseños muestrales, Muestreo probabilístico, Aleatorio Simple, Muestreo Probabilístico estratificado, Estimaciones, Variabilidad, Coeficientes de variación.

## Summary

As is known, satisfaction and perception are the most important indicators to measure quality of services, it becomes necessary to apply, after two years of implementing the district's financing program to productive units owned by vulnerable and special population mainly, a survey of satisfaction and perception with the objective of knowing the evaluation and opinions of the beneficiary citizens about the program. Due to the approach of the Colombian Government, this research seeks to compare different sample designs to determine the most representative sample of the database of beneficiaries with credits for the strengthening of their productive units in the city of Bogota.

**Keywords:** Finite Populations, Categorical Variables, Sample Designs, Probabilistic Sampling, Simple Random, Stratified Probabilistic Sampling, Estimates, Variability, Variation Coefficients

## Introducción

El proyecto establece la priorización de financiamiento a grupos poblaciones especiales y vulnerables como: desplazados, reinsertados, mujeres, mujeres cabeza de familia, discapacitados jóvenes, adultos mayores y grupos étnicos, entre otros. Motivo por el cual el marco muestral con el que se trabajará, contiene estas variables poblacionales, que serán descritas en la primera sección de este trabajo de grado.

Es por ello que el caso de la población que estos modelos estadísticos se aplicarán para proyectos de interés social beneficiarios de créditos para el fortalecimiento de sus unidades productivas en la ciudad de Bogotá, de allí se realizara la verificación de las herramientas y los resultados que arrojen y de esta manera verificar cual es la más óptima y que permita el mejor análisis de los datos.

El objetivo general de la investigación busca comparar diferentes diseños muestrales para determinar aquél que encuentra la muestra más representativa de la base de datos de los beneficiarios con créditos para el fortalecimiento de sus unidades productivas en la ciudad de Bogotá.

Donde se resalta la implementación de procesos estadísticos basados en la selección del diseño para encuestas muestrales en proyectos de interés socioeconómico en el marco de aplicación a beneficiarios de créditos para el fortalecimiento de sus unidades productivas en la ciudad de Bogotá D.C.

Para seleccionar el mejor diseño muestral no es suficiente comparar los coeficientes de variación obtenidos para la variable MONTO DESEMBOLSADO ya que todos los diseños arrojan valores muy cercanos al interior de las categorías y de cero para el total de la estimación de la variable; a su vez al utilizar un muestreo estratificado es imprescindible garantizar que el diseño asegure que se seleccione al menos una unidad de muestreo de cada estrato. Al momento de seleccionar el diseño es importante considerar la intención del muestreo, por ejemplo si es para realizar encuestas, los costos de aplicación de las mismas resultan de gran importancia por lo que podría pensarse en seleccionar previamente conglomerados. Dentro de las conclusiones se evidencia que las unidades de muestreo al interior de cada conglomerado deben ser lo más heterogéneas posible. Para el presente estudio el diseño muestral más adecuado resulta ser el bietápico por conglomerados y estratificado por la variable SEXO no solo por presentar muy poca variabilidad (coeficiente de variación) al interior de las clases de la variable MONTO DESEMBOLSADO si no que el hecho de utilizar conglomerados reduce los costos de aplicación de las encuestas y garantiza representatividad de toda la población.

Como autores se recomienda lograr estudios de mayor impacto académico, con una mayor muestra, por medio del cual se logra dar continuidad a la aplicabilidad de los métodos estadísticos, priorizando su efectividad.

## Objeto de estudio

### Presentación del problema

Uno de los principales obstáculos para el desarrollo de nuevas empresas dinámicas, particularmente en América Latina, es el financiamiento. Los emprendedores de esta región tienen menor acceso a fuentes especializadas de financiamiento y Bogotá no es la excepción, como consecuencia se crean menos empresas y las que logran nacer lo hacen con un tamaño y nivel tecnológico que no es el óptimo (Niño Riaño & Velandia, 2011).

Por lo anterior el Distrito optó por implementar dentro de su plan de desarrollo 2012-2016, un programa de inversión con el cual las unidades productivas de la economía popular de Bogotá D.C. puedan tener mayor acceso a recursos financieros (Concejo de Bogotá, 2012).

En razón a que a la fecha se han alcanzado 1804 beneficiarios, la entidad a cargo del proyecto ha optado por tomar una muestra en lugar de realizar un censo al total de los beneficiarios, para la aplicación de las encuestas de satisfacción y percepción. Lo anterior se debe principalmente a tres razones:

Un censo completo resulta más costoso por cuestiones de papelería, personal, desplazamiento etc. Trabajar con una muestra representativa permitirá tener resultados en menor tiempo, esto tanto por el acceso a toda la población que podría no estar disponible en el momento para diligenciar la encuesta, como por el procesamiento de los datos.

El censo puede resultar menos exacto que la muestra. Esto debido a que mayores volúmenes de información generan en algunos casos errores en la toma de datos, digitación y otros tipos de errores propios de los censos. Con la presente investigación se pretende establecer procedimientos metodológicos para la construcción del diseño muestral para encuestas, buscando lograr un muestreo eficiente y económico que permita hacer conclusiones e inferencia a toda la población; por tanto las técnicas utilizadas estarán enmarcadas dentro del muestreo probabilístico. Bajo la importancia de la estadística se hace indispensable cuantificar y cualificar la información que se tiene en el entorno para así lograr obtener diversos resultados de una manera más simple de comprender y analizar en el momento de comparar datos, es por ello que el ser humano ha utilizado la estadística y sus diferentes herramientas para poder tener resultados más comprensibles en bases de datos grandes y lograr obtener información basados en datos numéricos o con un patrón que permita cuantificar y cualificar la información que se posee (Pacheco Espejel, 2013).

Con el presente trabajo se pretende utilizar tres modelos estadísticos como lo son el modelo aleatorio simple, el estratificado proporcionado y el bietápico para determinar la implementación del estudio de un proyecto de interés social tomando como muestra la población de beneficiarios de créditos para el fortalecimiento de sus unidades productivas en la ciudad de Bogotá, con lo cual se pretende definir cuál de estas técnicas permitirá obtener la información más clara y concisa de esta población a estudiar, así como también, la utilización de los recursos, número de emprendedores, niveles de éxito de las empresas, tipo de población que recibió los recursos (Cámara de Comercio de Bogotá, 2012).

La importancia de la utilización de estos modelos estadísticos es definir cuál de las herramientas a utilizar pueden brindar una mejor información de los datos que se pretenden tabular porque la población a trabajar es población vulnerable bajos recursos, a las cuales el estado les brinda diferentes mecanismos de apalancamiento financiero y de capacitación, por medio de la cual logren emprender su propio negocio, es por ello que se debe conocer la utilización de los recursos, de igual forma sus niveles de éxito en sus actividades comerciales o de servicios, es de vital importancia reconocer que el mejoramiento de la calidad de vida de estas personas y la generación de empleo que ellos contribuyen se verá reflejada en beneficios propios y para la comunidad donde habitan.

Lo anterior se concreta en la pregunta de investigación: ¿Cuál es el diseño muestral más idóneo para encontrar la muestra más representativa de la base de datos de los beneficiarios con créditos para el fortalecimiento de sus unidades productivas en la ciudad de Bogotá?

### **Justificación del problema**

El muestro aleatorio simple es un procedimiento monoetápico que se puede calcular con o sin reposición, básicamente la diferenciación se basa en la dependencia o independencia de la selección de los datos que se tomen en cuenta a considerar ya que se extrae una porción finita de la población de los beneficiarios de créditos de fortalecimiento de las unidades productivas para la ciudad de Bogotá, se toma una muestra al azar dado que la población es demasiado grande se debe limitar ya que cada elemento que se toma tiene el mismo valor o peso al momento de aplicar los cálculos los cuales proporcionen los datos que permitan obtener información deseada que brinde amplios puntos de comparación. Esto a raíz de la variabilidad de los sub grupos que se hacen ya que es menor la comparación de variables, esta técnica es la más simple de todas ya que se realiza de manera intuitiva y sus cálculos son muy sencillos los resultados que arroja hacen que sea muy exacta y tiene altos niveles de precisión lo cual genera menor operatividad y menores costos, tiempo y esfuerzo en la aplicación de esta técnica (Vivanco, 2005).

Por otra parte el muestreo estratificado proporcionado se utiliza cuando la población se puede subdividir. Esta técnica es más utilizada en grupos homogéneos a los que se les llama estratos, esta clasificación puede ser por las diferentes variables que presente la muestra (género, edad, profesión, ubicación geográfica etc.), y permite ir más al detalle del estudio de la población, permite llegar a una estratificación que en ocasiones parece inaccesible lo que permite llegar a realizar pruebas donde la población es más difícil de evaluar, ya que arroja datos más exactos acerca de la muestra en cuanto a la estratificación del grupo, esta muestra de la población debe ser proporcional al tamaño de la misma y la estratificación que se realice (Malhotra, 2004).

Así como también encontramos el muestreo bietápico este es el cual se hacen muestreos aleatorios en dos fases en la primera se realiza una selección por conglomerados y es desde allí donde se hace la selección de la muestra y la segunda fase es cuando se realiza el muestreo de las unidades que lo conformaran (Pérez Rigoberto, 2010).

Dado que la población del presente caso de estudio está conformada por los beneficiarios de un proyecto de interés social, donde son de esencial importancia las características particulares de los individuos de la población, se hace necesario que la muestra seleccionada, para futuros estudios de satisfacción, sea representativa de la población es decir que conserve en proporción la diversidad de particularidades que se presentan. La estratificación y los modelos complejos de muestreo pueden garantizar que se cumplan estas condiciones al momento de seleccionar la muestra.

## **Objetivo de la investigación**

### **Objetivo general.**

Comparar diferentes diseños muestrales para determinar aquel que encuentra la muestra más representativa de la base de datos de los beneficiarios con créditos para el fortalecimiento de sus unidades productivas en la ciudad de Bogotá.

### **Objetivos específicos.**

- Analizar el comportamiento de las variables categóricas para determinar su interés o relevancia en la construcción de los diseños muestrales.
- Encontrar relaciones de dependencia entre las variables explicativas y la variable de dependiente MONTO DESEMBOLSADO.
- Medir la eficiencia del M.A.S. Para el caso de estudio mediante el cálculo del coeficiente de variación al interior de las categorías de la variable objetivo y las variables explicativas.
- Medir la eficiencia del Muestreo Estratificado. Para el caso de estudio mediante el cálculo del coeficiente de variación al interior de las categorías de la variable objetivo y las variables explicativas.
- Medir la eficiencia de diseños muestrales complejos. Para el caso de estudio mediante el cálculo del coeficiente de variación al interior de las categorías de la variable objetivo y las variables explicativas.
- Comparar la eficiencia de los diseños muestrales para determinar cuál permite obtener la muestra más representativa de la población.

## **Marco teórico**

### **Antecedentes de la investigación**

Diferentes organizaciones en el mundo como ONU, CEPAL, OCDE y en Colombia DANE, ANDI, que trabajan en asuntos económicos y sociales utilizan las encuestas como fuente de datos para estadísticas sociales y demográficas. Es fácil conseguir información sobre como planear, diseñar y aplicar encuestas, pero no es mucha la información sobre diseños muestrales(Sánchez Pérez, Duarte, & Blanco, 2013).

Revisando exhaustivamente se encuentra que prácticamente todos los diseños muestrales para encuestas asociadas a estudios socioeconómicos son complejos debido a sus características: varias etapas, estratificación y conglomeración. Además, el hecho de que con frecuencia las encuestas sean a escala nacional y suelen tener un alcance de carácter general, por cubrir varios temas de interés para organizaciones gubernamentales, añade complejidad(Misas Gabriel, 2004).

En Bogotá entidades como el IPES (Instituto para la economía social) es la encargada de crear políticas de estructura organización y funcionamiento la cual ejecuta programas en base a los planes de desarrollo de los gobiernos distritales que le permite a sectores tales como la economía informal, a través de programas de capacitación de capital humano y posterior acceso a créditos que permiten a esta población a la inserción en los mercados de la economía del país y la reinserción de estas personas al comercio o prestación de servicios de manera formal(Morera Rojas, 2013).

Con la presente investigación se pretende comparar tres diseños muestrales generalmente utilizados en encuestas de interés socioeconómico y con el fin de determinar cual resulta más adecuado para aplicar encuestas de satisfacción a los beneficiarios de créditos para el fortalecimiento de sus unidades productivas en la ciudad de Bogotá. Los tres diseños seleccionados están basados en los modelos estadísticos probabilísticos, los cuales parten de modelos matemáticos donde se reúne un conjunto de distribuciones de probabilidad. Se escogen estos ya que se supone son los que pueden presentar datos más pertinentes para los futuros análisis (Cárdenas Santa María Mauricio, 2013).

### **Análisis de variables categóricas**

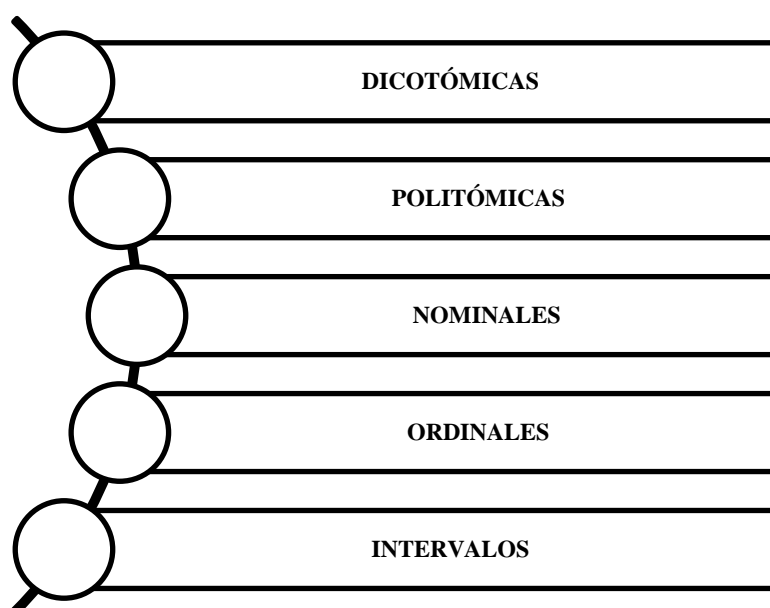
Mediante el análisis de datos bajo los modelos estadísticos aleatorio simple, estratificado proporcionado y el modelo bietápico se evaluarán las variables categóricas y cualitativas identificando aquellas más relevantes que permitan clasificar o generar un patrón homogéneo de criterios relevantes con los cuales se pueda generar el mismo estudio y de esta manera se pueda realizar un paralelo comparativo que permitan cuantificar el grado de asociación y mayor veracidad de la información que se tiene(Berenson & Levine, 1996).



Dado que la población de estudio se encuentra en la ciudad de Bogotá, beneficiarios de créditos para fortalecimiento de sus unidades productivas, y que por las particularidades del programa la población es heterogénea, por lo cual se debe ir más al detalle de la misma e indagar en variables como género, estrato socioeconómico, estado civil, nivel de escolaridad etc., para clasificar la información de manera más exhaustiva.

En el protocolo de variables de un estudio estadístico se tienen en cuenta todas aquellas que son cuantificables y cualitativas. Es por ello que los datos que recolectan o se tienen en cuenta hacen objeto de la investigación, de aquí la importancia de clasificar la información que se va a tabular y generar una correlación entre los muestreos a utilizar de esta manera.

**Figura 1** Clasificación de las variables estadísticas.



Tipos de variables. Tomado de : título, (Sabadías, 1995), modificado por los autores, 2017.

Variables dicotómicas son de tipo cualitativo y solo existen dos variables a escoger como por ejemplo género hombre o Mujer en general estos resultados son tabulados por una operación binaria (Pineda, 1998).

**Tabla 1** Estadística descriptiva e inferencial.

<b>Sexo</b>	1	0	0	1	0	1	1	1	0	0	1	0	1	1	0	1
<b>Puntuación</b>	24	17	8	31	24	16	25	30	31	17	14	16	20	8	16	24

Tomado de: Variables Dicotómicas Tomado de: Estadística descriptiva e inferencial.  
Universidad de Castilla La Mancha, (Sabadías, 1995). Modificado por los autores 2017.

En cuanto a las variables polinómicas, también son de tipo cualitativo y son el resultado de una respuesta múltiple como lo sería un lugar de nacimiento, estrato social un color entre otras en estas hay más de dos variables a considerar (García, 1990).

En comparación, la nominal es una clasificación un poco más amplia ya que puede ser concerniente a respuestas como a que religión, tipo de raza o partido político pertenece.

**Tabla 2** Tipos de variables.

Población	Cuantitativas		Categorías	
	Continuas	Discretas	Ordinales	Nominales
Animales de laboratorio	Peso	Número de crías	Calidad del nido: Excelente, bueno, regular, malo	Color: Negro, blanco, gris
	Edad			
Automóviles	Peso	Número de defectos por auto	Tamaño: Compacto, mediano, grande  Precio: Barato, medio, costoso	Colores, marca, fabricante
	Longitud			

Tomado de: UNAM, (Márquez María, 2009). Modificado por los autores 2017.

Con el método ordinal es posible establecer un orden de categorías ya que se puede referir a un rango, tipo de clase social pero es un poco sesgada ya que no se puede definir a una persona a que clase social pertenece ya que los poderes adquisitivos pueden variar entre los individuos en esta se asignan los valores de manera aleatoria por lo que no poseen un orden lógico no son operables matemáticamente (Alvarado Jorge, 2008).

La variable por intervalos procede a variables cuantitativas que presentan un intervalo como su nombre lo indica como por ejemplo un rango salarial, o un rango entre distancias estas usualmente son ordenadas por rangos (Ruiz, 1993).

### **Tablas de contingencia.**

En el análisis de variables categóricas se acostumbra organizar los datos en tablas de doble entrada en las que cada entrada representa un criterio de clasificación. Como resultado de esta clasificación, las frecuencias aparecen organizadas en casillas que contienen información sobre la relación existente entre ambos criterios (Turriago, 2011).

La tabla 3 muestra un ejemplo de tabla de contingencia. En ella 300 sujetos han sido ordenados con arreglo a dos criterios: sexo y estrato socioeconómico. Los números que

aparecen en la tabla son frecuencias absolutas (número de casos): 13 hombres pertenecen al estrato 1, 18 mujeres pertenecen al estrato 1, 55 mujeres pertenecen al estrato 4, etc.

**Tabla 3** Tabla de contingencia para las variables sexo y estrato socioeconómico.

		Estrato socio-económico					Total
		Estrato 1	Estrato 2	Estrato 3	Estrato 4	Estrato 5	
Sexo	Hombre	13	56	50	1	1	121
	Mujer	18	101	55	4	1	179
Total		31	157	105	5	2	300

Fuente los autores.

### Estimador Chi-cuadrado de Pearson.

El grado de relación entre dos variables categóricas no puede ser establecido simplemente observando las frecuencias en una tabla de contingencia. La simple observación de las frecuencias no permite una conclusión definitiva. Para determinar si dos variables se encuentran relacionadas se debe utilizar alguna medida de asociación acompañada de su correspondiente prueba de significación.

El estadístico *Chi-cuadrado* propuesto por Pearson (1911) permite contrastar la hipótesis de que los dos criterios de clasificación utilizados (variables categóricas) son independientes: Para ello, compara las frecuencias observadas con las frecuencias esperadas, estas son las frecuencias que teóricamente se encontrarían en cada casilla si los dos criterios fueran independientes (Peña, 2008).

El estadístico Chi-cuadrado se construye a partir de la distribución de Pearson o ji-cuadrada.

En estadística, la distribución de Pearson, llamada también ji cuadrada(o) o chi cuadrado(a) ( $\chi^2$ ), es una distribución de probabilidad continua con un parámetro que representa los grados de libertad de la variable aleatoria. Donde son variables aleatorias normales, independientes de media cero y varianza uno. El que la variable aleatoria tenga esta distribución se representa habitualmente así:

### Formula 1 Estadístico *Chi-cuadrado*.

$$\chi^2 = \sum_i \frac{observada_i - teorica_i}{teorica_i}^2$$

Tomado de: Fundamentos de estadística (Peña, 2008).

Cuanto mayor sea el valor, menos verosímil es que la hipótesis nula (que asume la igualdad entre ambas distribuciones) sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de chi-cuadrado, más ajustadas están ambas distribuciones.

Los grados de libertad  $gl$  vienen dados por  $r$  y  $k$ , donde  $r$  es el número de filas y  $k$  el de columnas.

- Criterio de decisión:

No se rechaza cuando. En caso contrario sí se rechaza. Donde  $t$  representa el valor proporcionado por las tablas, según el nivel de significación estadística elegido.

### **Diseños muestrales.**

El muestreo es un procedimiento que responde a la necesidad de información estadística precisa sobre la población y los conjuntos de elementos que la conforman; el muestreo trata con investigaciones parciales sobre la población que apuntan a inferir a la población completa (Levin & Rubin, 2004).

El muestro se aplica en los casos en que no es óptimo o viable realizar un censo. Es decir, encuestar al total de las unidades que conforman una población, es ahí donde debe seleccionarse una muestra, la cual debe corresponder a una sección representativa de la población, de manera que sea posible hacer inferencia de los resultados obtenidos en la muestra, sobre toda la población. Sin embargo en la vida real se presentan muchos casos en los que no es posible conocer la población, es decir no se cuenta con un marco muestral, por lo que la selección de la muestra no resulta representativa de una población, y por tanto los resultados no se pueden extrapolar. Los casos en que se cuenta con un marco muestral, es posible hacer un muestreo probabilístico, el cual debe cumplir con dos requisitos: Todos los elementos de la población tienen la misma probabilidad de ser seleccionados y esta probabilidad de ser seleccionados en la muestra es mayor a cero (Berenson & Levine, 1996).

Los métodos de muestreo probabilísticos más utilizados son: Muestreo Aleatorio Simple, Muestreo Aleatorio Sistemático, Muestreo Aleatorio Estratificado y Muestreo Aleatorio por Conglomerados. Cuando la selección de la muestra no se realiza aleatoriamente, sino a través de un procedimiento subjetivo o a criterio del experto o investigador, de manera que la probabilidad de ser seleccionados en la muestra no es igual para todos los elementos de la población, se conoce como muestro no probabilístico (Salkind, 1999).

Dado que todos los elementos del marco muestral del presente trabajo tienen la misma probabilidad de ser seleccionados, se utilizarán diseños muestrales probabilísticos como el muestreo aleatorio simple, el muestreo estratificado y el muestreo bietápico por conglomerados y estratificado con el fin de establecer cuál de los tres diseños presenta mejores resultados.

### **Muestreo Aleatorio Simple.**

La técnica de muestreo aleatorio simple es la técnica de muestreo más sencilla, donde cada miembro de la población tiene igual probabilidad de ser seleccionado, lo que la hace una técnica fácil y justa. La dificultad que normalmente se presenta para su aplicación está relacionada con la exigencia de contar con la lista completa de todos los miembros de la población (Scheaffer, Mendenhall, & Ott, 2006).

**Formula 2** Muestreo Aleatorio Simple.

$$n = \frac{n'}{1 + \frac{n'}{N}}$$

Tomado de: Elementos de muestreo, (Scheaffer et al., 2006).

### **Muestreo Probabilístico Estratificado.**

La técnica de muestreo probabilístico estratificado es aquella donde se divide la población en subgrupos o estratos, de estos grupos o estratos se selecciona una muestra aleatoria, tomando al menos un elemento de cada grupo o estrato (Abascal & Esteban, 2005).

Al formar los estratos se debe buscar que los elementos de cada estrato sean lo más homogéneos entre sí y lo más heterogéneos entre los estratos, lo que genera menor varianza de las estimaciones. Estos estratos pueden reflejar regiones geográficas de un país, clases sociales dentro de una ciudad, etc.

### **Muestreo Estratificado Proporcional.**

Es una técnica de carácter probabilístico en donde la persona que realiza la investigación hace una clasificación de la información realizando una subdivisión de la población o la estratifica como su nombre lo indica, posteriormente realiza una selección aleatoria de cada subdivisión y toma una muestra aleatoria de cada estrato que haya asignado debe ser de forma proporcional, esto se hace a razón de que la muestra puede estar desproporcionada por eso se clasifican los individuos según interés de la investigación (Vivanco, 2005).

**Formula 3** Muestreo Estratificado Proporcional.

$$\frac{Z^2 \sum_{h=1}^H W_h P_h (1 - P_h)}{e^2}$$

Tomado de: Muestreo Estadístico. Diseño Y Aplicaciones, (Vivanco, 2005).

### **Muestreo por Conglomerados.**

Es una técnica que se usa cuando hay homogeneidad en la muestra de manera natural y está compuesta por dos o más elementos de la población a investigar, cabe destacar que la inferencia no se efectúa a partir de la observación directa de las variables que la comprenden, se hace una selección de la muestra la cual se clasifica en estos grupos también llamados clusters, se obtiene la información de una muestra aleatoria simple de los elementos que se seleccionan de cada grupo posteriormente se toma una submuestra de los mismos esta se utiliza cuando se quiere reducir el número de encuestas a realizar en la población. También ayuda a reducir costos, esto influye en la precisión de la información que se requiera. Es importante destacar que existen conglomerados con submuestreo y sin submuestreo se hace relevante hacer esta diferenciación ya que esta da pie al muestreo por etapas (Freund & Simon, 1994).

### **Muestreo en etapas.**

Seleccionar la muestra en etapas tiene ventajas prácticas para el proceso de selección mismo. Permite al muestreador aislar en pasos sucesivos los lugares geográficos donde se llevarán a cabo las actividades de la encuesta. Con el muestreo por conglomerados hay un mínimo de dos etapas en el procedimiento de selección: la primera corresponde a la selección de los conglomerados, y la segunda a la selección de las unidades de muestreo. Los conglomerados de las encuestas socioeconómicas generalmente se definen como unidades geográficas de algún tipo. En la segunda etapa se puede aplicar M.A.S o muestreo estratificado (Moore, 2005).

### **Muestreo Complejo.**

Generalmente los diseños muestrales empleados en las encuestas socioeconómicas se desvían del muestreo aleatorio simple por presentar una o más de las siguientes características:

1. La estratificación en una o más etapas de muestreo;
2. La conglomeración de las unidades en una o más etapas de muestreo
3. La ponderación para compensar imperfecciones de la muestra tales como las probabilidades desiguales de selección, la falta de respuesta y la falta de cobertura.

Un diseño muestral se denomina complejo cuando en él se da una o más de las características anteriores (Kendall & Kendall, 1997).

### **Eficiencia del Diseño Muestral**

El análisis y evaluación de los diseños muestrales es de particular importancia en las grandes encuestas de interés socioeconómico, que generalmente se diseñan para que sean la fuente única de información precisa en gran variedad de temas asociados. La medida clave de precisión en las encuestas por muestreo es la varianza de muestreo, un indicador de la

variabilidad que se introduce al elegir una muestra asumiendo que la información recopilada en la encuesta es veraz. La varianza de muestreo es una medida de la variabilidad de la distribución de muestreo de un estimador. El error estándar, o raíz cuadrada de la varianza, se emplea para medir el error de muestreo. Un estimador de este error de muestreo puede emplearse para indicar la exactitud de las estimaciones (Hernández, Cantin Garcia, Lopez Abejon, & Rodriguez Zazo, 2014).

La forma del estimador de la varianza y el modo como se evalúa dependen del diseño muestral en que este basado. En los diseños estándar estos estimadores se suelen evaluar mediante el uso de fórmulas simples. Sin embargo, en los diseños muestrales complejos utilizados en las encuestas de interés socioeconómico que por lo general implican el uso de estratificación, conglomerados y muestreo con probabilidad desigual, las formas de estos estimadores suelen ser complejas y difíciles de evaluar. El cálculo de los errores de muestreo en este caso requiere procedimientos que tengan en cuenta la complejidad del diseño muestral con el que se generaron los datos, que a su vez requiere el uso de programas informáticos apropiados (Azorín, Poch, & Sánchez-Crespo, 1986).

Además de la varianza de muestreo hay otras formas de medir el error de muestreo, como el error estándar, el coeficiente de variación y el efecto del diseño, que guardan entre sí una relación algebraica, en el sentido de que es posible deducir la expresión de cualquiera de ellas a partir de las otras mediante la aplicación de sencillas operaciones algebraicas. En la presente investigación estas medidas del error de muestreo se calcularán con ayuda del paquete estadístico SPSS por lo cual no son de interés las complejas formas de calcularlos, más si lo es su interpretación (Taborga, Castellón, & Taborga, 2012).

**El error estándar** de un estimador es la raíz cuadrada de su varianza de muestreo. Esta medida resulta más fácil de interpretar porque para la indicación del error de muestreo utiliza la misma escala que la estimación, mientras que la varianza se basa en diferencias cuadráticas (Scheaffer et al., 2006).

**Formula 4** Error estándar.

$$SE_x = \frac{S}{n}$$

Tomado de: Elementos de muestreo, (Scheaffer et al., 2006).

**El coeficiente de variación (CV)** de una estimación es el coeficiente entre su error estándar y el valor promedio de la propia estimación. Así, el coeficiente de variación proporciona una medida del error estándar respecto a la característica que se está midiendo. Suele expresarse en porcentajes. El coeficiente de variación es útil para comparar la precisión de estimaciones de una encuesta cuyo tamaño o escala difieren. Sin embargo, no resulta útil para estimadores de características cuyo valor real pueda ser cero o negativo, incluidas

estimaciones de cambio como, por ejemplo, el cambio en los ingresos promedio a lo largo de dos años (Vázquez & Ortiz, 2005).

**Formula 5** Coeficiente de variación.

$$C_v = \frac{\vartheta}{x}$$

Tomado de: Métodos estadísticos para medir, describir y controlar la variabilidad (Vázquez & Ortiz, 2005).

**El efecto del diseño (deff)** se define como el coeficiente entre la varianza de muestreo de un estimador en un diseño determinado y la varianza de muestreo del estimador basada en una muestra aleatoria simple del mismo tamaño. Se considera también el factor por el cual hay que multiplicar la varianza de una estimación basada en una muestra aleatoria simple del mismo tamaño para tener en cuenta las complejidades del diseño muestral derivadas de factores como la estratificación, la conglomeración o las ponderaciones. (Brogan, 2005).

**Formula 6** Formula del efecto del diseño (deff).

$$Deff = M \frac{\sum_{i=1}^{i=M} p_i - P}{NPQ}$$

Tomado de: Muestreo por conglomerados en encuestas poblacionales

Cluster sampling in health surveys (Zapata-Ossa et al., 2011).



## **Metodología de la investigación**

### **Tipo de estudio**

La presente investigación es de carácter Descriptivo – Correlacional ya que en la primera parte se realiza un análisis descriptivo de las variables y se buscan relaciones entre ellas, además se seleccionan tres diseños muestrales con el fin de determinar cuál de ellos permite seleccionar la mejor muestra (la que mejor representa a la población).

### **Población objeto**

Beneficiarios de créditos para el fortalecimiento de sus unidades productivas en la ciudad de Bogotá y municipios aledaños (Grupo de individuos de los que se obtiene información).

### **Unidades de análisis**

Unidad Productiva. (Individuo que se investiga).

### **Unidad de muestreo**

Beneficiarios crédito. (Unidad de la población a partir de las cuales se selecciona la muestra).

### **Marco muestral**

Listado de beneficiarios de créditos para el fortalecimiento de sus Unidades Productivas o proyectos de emprendimiento ubicados en Bogotá y municipios aledaños

### **Variables**

Las tablas 4 y 5 presentan la descripción de las variables de estudio, en adelante se utilizarán indistintamente el nombre o la abreviatura de las variables.

**Tabla 4** Descripción de las variables, parte 1.

Nombre variable	Abreviatura	Descripción	Rango de valores	Tipo de variable
Unidad productiva	TEUP	Tipo de emprendimiento o unidad productiva	Tipo 1: idea de emprendimiento  Tipo 2: unidad productiva  Tipo 3: asociación	Cualitativa Nominal
Sector económico	SEEC	Sector económico al que pertenece la unidad productiva	Agropecuario  Comercio  Industria  Servicios	Cualitativa Nominal
Organización jurídica	ORJU	Tipo de organización jurídica en que se ubica la unidad productiva	Asociación  Corporación  Economía solidaria  Fami - empresa  Fundación  unipersonal  Ltda.  Ong.  Persona natural  S.A.  Sin ánimo de lucro	Cualitativa Nominal

Fuente los autores, 2017.

**Tabla 5** Descripción de las variables, parte 2.

<b>Nombre variable</b>	<b>Abreviatura</b>	<b>Descripción</b>	<b>Rango de valores</b>	<b>Tipo de variable</b>
Número trabajadores	NUTR	Número de trabajadores con que cuenta la unidad productiva	0 a 100	Cuantitativa discreta
Localidad	LOMU	Localidad geográfica donde se ubica la unidad productiva	18 localidades Bogotá encontradas en la base de datos.	Cualitativa Nominal
Rango de edad	EDAD	Rango de edad del representante de la unidad productiva	Adulto (27 - 59 años) Joven (14 - 26 años) Mayor (mayor de 60 años)	Cualitativa Nominal
Sexo	SEXO	Sexo del representante legal de la unidad productiva	Hombre Mujer	Cualitativa Nominal
Nivel educativo	NIED	Nivel educativo del representante legal de la unidad productiva	Ninguno Postgrado completa Postgrado incompleta Primaria completa Primaria incompleta Secundaria incompleta Secundaria completa Técnica completa Técnica/tecnológica completa Técnica/tecnológica incompleta Universitaria	Cualitativa Nominal

			completa	
			Universitaria incompleta	
Estrato socioeconómico	ESEC	Estrato socioeconómico del representante legal de la unidad productiva	Estrato 1 Estrato 2 Estrato 3 Estrato 4 Estrato 5	Cualitativa Nominal
Monto desembolsado	MODE	Monto del crédito desembolsado para el fortalecimiento de la unidad productiva	\$0 hasta \$100 millones	Cuantitativa discreta

Fuente los autores, 2017.

### **Análisis descriptivo de las variables**

Para determinar las características de interés de las variables de estudio se realizarán los siguientes análisis para cada una de ellas:

- Se realiza análisis por tablas de frecuencias para las variables nominales.
- Se determina si es necesario agrupar categorías en algunas de ellas por presentar elevado número de clases.

### **Relaciones de dependencia**

Se realiza un análisis del Monto desembolsado categorizado por diferentes variables para determinar si existen relaciones de dependencia. Luego entre las variables explicativas se realizan pruebas de independencia para reducir su número de ser posible.

### **Selección del diseño muestral**

Revisando antecedentes se observa que el Diseño Muestral más apropiado para encuestas a hogares, empresas o comercio debe ser probabilístico, ya que es necesario generalizar los resultados a toda la población. A su vez polietápico, estratificado y por conglomerados.

Por lo anterior se propone comparar tres diseños cada uno más complejo que el anterior:

- a) Muestreo Aleatorio Simple
- b) Muestreo Estratificado
- c) Muestreo bietápico por conglomerados y estratificado

Es necesario elaborar tablas de contingencia para determinar si las variables explicativas están relacionadas. Una vez se determine cuáles variables son independientes se utilizan estas como variables de estratificación en los diseños muestrales complejos. También se utilizará la localidad (una vez se agrupe) como variable para elaborar conglomerados.

### **Cálculo del tamaño de muestra**

Dado que la intención de la investigación es comparar diseños muestrales para determinar cuál es más eficiente, se determinará un tamaño acorde a cada uno de los diseños planteados.

### **Selección de la muestra**

Se realizan planes muestrales para los tres diseños propuestos:

- Para el caso del M.A.S. solamente se requiere previamente calcular el tamaño de muestra.
- Para el caso del Muestreo Estratificado es necesario determinar previamente cual va ser la variable de estratificación y calcular el tamaño de muestra para cada estrato.
- Para el caso del Muestreo Bietápico es necesario determinar cómo se van a formar los conglomerados y determinar la variable de estratificación, así como calcular el tamaño de muestra para cada una de las etapas.

### **Eficiencia del diseño muestral**

Para determinar la eficiencia del diseño muestral se compara la capacidad de predecir el tamaño de población categorizado por diferentes variables explicativas, además se consideran el Coeficiente de Variación y el Efecto del diseño.

## Resultados de la investigación

### Análisis descriptivo

Se construyeron tablas de contingencia y diagramas de barras (Ver Apéndice 1) para analizar el comportamiento de las 8 variables categóricas obteniendo los siguientes resultados de interés para la investigación:

#### Tipo de emprendimiento o unidad productiva.

Esta variable categoriza la población en tres grupos (tipos de emprendimiento) encontrando que el 88% de los beneficiarios de crédito son unidades productivas por tanto no es una variable que represente importancia para el análisis del MONTO DESEMBOLSADO o para la selección de la muestra. Ver Tabla 6.

**Tabla 6** Tabla de frecuencias para la variable TEUP.

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Tipo 1: Idea de emprendimiento	215	11,9	11,9	11,9
Tipo 2: Unidad productiva	1587	88,0	88,0	99,9
Tipo 3: Asociación	2	0,1	0,1	100,0
Total	1804	100,0	100,0	

Fuente los autores, 2017.

#### Sector económico.

Esta variable posee cuatro categorías pero dado que Bogotá es en su mayoría urbana, menos de un 1% de los beneficiados pertenecen al sector agropecuario. Resulta conveniente incluir el sector agropecuario dentro del sector industrial. La Tabla 7 presenta las frecuencias para la variable SECTOR ECONÓMICO.

**Tabla 7** Tabla de frecuencia para la variable SEEC.

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Comercio	644	35,7	35,7	35,7
Industria	755	41,9	41,9	77,5
Servicios	405	22,5	22,5	100,0
Total	1804	100,0	100,0	

Fuente los autores, 2017.

### Organización jurídica.

El 91% de los beneficiados con créditos son personas jurídicas mientras que las once categorías restantes apenas constituyen un 9%, por tanto no es una variable que represente importancia para el análisis del MONTO DESEMBOLSADO o para la selección de la muestra. Ver Apéndice A.

### Número de trabajadores.

Dado que el 98,4% de las unidades productivas tienen 10 o menos empleados, se hace necesario reagrupar para obtener menor número de clases. La tabla 8 presenta la variable NUTR agrupada en 5 clases.

**Tabla 8** Tabla de frecuencias para la variable NUTR (agrupada).

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<= 1,00	608	33,7	33,7	33,7
2,00 - 2,00	550	30,5	30,5	64,2
3,00 - 3,00	269	14,9	14,9	79,1
4,00 - 4,00	173	9,6	9,6	88,7
5,00+	204	11,3	11,3	100,0
Total	1804	100,0	100,0	

Fuente los autores, 2017.

Agrupando en 5 clases se observa que más del 60% de las unidades de negocio tienen a lo sumo dos trabajadores y que apenas un 11% tienen cinco o más empleados.

### Localidad/Municipio.

Esta variable presenta 24 categorías de las cuales Kennedy es la de mayor representación con el 21,9% de los beneficiados y dado que esta variable es requerida para seleccionar los conglomerados en el muestreo complejo, resulta necesario agrupar localidades para lo cual es importante realizarlo geográficamente por cuestión de costos.

Se agrupan las localidades y municipios aledaños en 5 zonas geográficas. La tabla 9 muestra las frecuencias absolutas y relativas para la variable ZONA GEOGRÁFICA.

**Tabla 9** Tabla de frecuencias para la variable ZONA GEOGRÁFICA.

<b>ZONA GEOGRÁFICA</b>	<b>Frecuencia</b>	<b>Porcentaje</b>	<b>Porcentaje válido</b>	<b>Porcentaje acumulado</b>
Centro	412	22,8	22,8	22,8
Norte	247	13,7	13,7	36,5
Occidente	581	32,2	32,2	68,7
Oriente	204	11,3	11,3	80,0
Sur	360	20,0	20,0	100,0
Total	1804	100,0	100,0	

Fuente los autores, 2017.

La variable ZONA GEOGRÁFICA será la nueva variable de conglomeración, presenta 5 categorías: centro, norte, occidente, oriente, sur.

### **Rango de edad.**

Esta variable concentra el mayor número de casos en la categoría ADULTO (27 a 59 años). La distribución de frecuencias para la variable se presenta en la tabla 10.

**Tabla 10** Tabla de frecuencias para variable EDAD.

<b>Rango de edad</b>	<b>Frecuencia</b>	<b>Porcentaje</b>	<b>Porcentaje válido</b>	<b>Porcentaje acumulado</b>
Adulto (27 - 59 años)	1582	87,7	87,7	87,7
Joven (14 - 26 años)	73	4,0	4,0	91,7
Mayor (mayor de 60 años)	149	8,3	8,3	100,0
Total	1804	100,0	100,0	

Fuente los autores, 2017.

Dado que la mayoría de casos (el 87,7%) pertenecen a una misma categoría (ADULTOS), es necesario revisar relaciones de dependencia con la variable MONTO DESEMBOLSADO para determinar si la variable es de utilidad para la selección de la muestra.



### Sexo.

**Tabla 11** Tabla de frecuencias para variable SEXO.

SEXO	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Hombre	787	43,6	43,6	43,6
Mujer	1017	56,4	56,4	100,0
Total	1804	100,0	100,0	

Fuente los autores, 2017.

La proporción de hombres beneficiados es del 43,6% frente a un 56,4% de mujeres beneficiadas, debido a la cercanía en las proporciones resulta ser una variable de interés al momento de estratificar.

### Nivel educativo.

La variable NIVEL EDUCATIVO distribuye la población en 13 clases, siendo algunas de ellas poco representativa (menos del 1%)

**Tabla 12** Tabla de frecuencias para variable NIVEL EDUCATIVO.

Nivel educativo	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Ninguno	23	1,3	1,3	1,3
Postgrado completa	16	,9	,9	2,2
Postgrado incompleta	4	,2	,2	2,4
Primaria completa	259	14,4	14,4	16,7
Primaria incompleta	73	4,0	4,0	20,8
Secundaria incompleta	238	13,2	13,2	34,0
Secundaria completa	628	34,8	34,8	68,8
Técnica completa	1	,1	,1	68,8
Técnica/tecnológica completa	291	16,1	16,1	85,0
Técnica/tecnológica incompleta	22	1,2	1,2	86,2
Universitaria completa	172	9,5	9,5	95,7
Universitaria incompleta	77	4,3	4,3	100,0
Total	1804	100,0	100,0	

Fuente los autores, 2017.

Demasiadas clases de ser requerida para la investigación debe reagruparse en menos categorías.

### Estrato socioeconómico.

La tabla 13 presenta la distribución de frecuencias para la variable estrato socioeconómico.

**Tabla 13** Tabla de frecuencias para variable ESTRATO SOCIOECONÓMICO.

<b>Estrato socioeconómico</b>	<b>Frecuencia</b>	<b>Porcentaje</b>	<b>Porcentaje Válido</b>	<b>Porcentaje acumulado</b>
Estrato 1	134	7,4	7,4	7,4
Estrato 2	988	54,8	54,8	62,2
Estrato 3	621	34,4	34,4	96,6
Estrato 4	50	2,8	2,8	99,4
Estrato 5	11	,6	,6	100,0
Total	1804	100,0	100,0	

Fuente los autores, 2017.

El 89,2% de los casos se encuentran en los estratos 2 y 3; en los estratos 1, 4 y 5 la frecuencia es muy baja por lo tanto no hay garantía que al seleccionar la muestra se seleccione por lo menos una unidad de cada estrato. Se decide NO utilizar la variable ESTRATO SOCIOECONÓMICO como variable de estratificación.

### Monto desembolsado.

Es la variable dependiente del presente estudio, es una variable cuantitativa discreta. Se realizan análisis de tendencia central y de variabilidad.

Existe mucha diferencia entre las medidas de tendencia central (ver tabla 14), además el Coeficiente de variación centrado en la media es de 70,9% lo que nos indica que existe alta variabilidad y por tanto la media no es representativa de la población.

**Tabla 14** Medidas de tendencia y variabilidad para la variable MONTO DESEMBOLSADO.

N	Válidos	1804
	Perdidos	0
Media		8412530,4490
Mediana		7000000,0000
Moda		5000000,00
Desv. típ.		5968508,98318
Varianza		35623099482259,210
Asimetría		3,599
Error típ. de asimetría		,058
Curtosis		34,440
Error típ. de curtosis		,115
Rango		95500000,00

Mínimo	500000,00
Suma	15176204930,00

Fuente los autores, 2017.

Debido a la alta variabilidad se decide categorizar la variable en cuatro clases como se muestra en la tabla 15.

**Tabla 15** Tabla de frecuencias para variable MONTO DESEMBOLSADO.

<b>Monto desembolsado Agrupado</b>	<b>Frecuencia</b>	<b>Porcentaje</b>	<b>Porcentaje válido</b>	<b>Porcentaje acumulado</b>
<= 5000000	730	40,5	40,5	40,5
5000001 - 11000000	693	38,4	38,4	78,9
11000001 - 17000000	272	15,1	15,1	94,0
17000001+	109	6,0	6,0	100,0
Total	1804	100,0	100,0	

Fuente los autores, 2017.

### **Relaciones de dependencia con la variable MONTO DESEMBOLSADO**

Se realiza un análisis entre la variable MONTO DESEMBOLSADO, por ser la variable de interés, y algunas de las variables explicativas mediante tablas de contingencia, diagramas de barras y el estadístico chi-cuadrado de Pearson para determinar si existen relaciones de dependencia.

#### **Monto desembolsado categorizado por sector económico.**

La tabla 16 presenta las pruebas de independencia para las variables SECTOR ECONÓMICO y MONTO DESEMBOLSADO.

**Tabla 16** Pruebas de independencia para las variables SEEC y MODE.

	<b>Valor</b>	<b>gl</b>	<b>Sig. asintótica (bilateral)</b>
Chi-cuadrado de Pearson	37,870 <sup>a</sup>	6	0,000
Razón de verosimilitudes	38,313	6	0,000
N de casos válidos	1804		
<i>Nota:</i> a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 24,47			

Fuente los autores, 2017.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables SEEC y MODE están relacionadas.

### **Monto desembolsado categorizado por número de trabajadores.**

La tabla 17 presenta las pruebas de independencia para las variables NUMERO TRABAJADORES y MONTO DESEMBOLSADO.

**Tabla 17** Pruebas de independencia para las variables NUTR y MODE.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	264,664 <sup>a</sup>	12	0,000
Razón de verosimilitudes	253,610	12	0,000
Asociación lineal por lineal	250,255	1	0,000
N de casos válidos	1804		
<i>Nota:</i> a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 10,45.			

Fuente los autores, 2017.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables NUTR y MODE están relacionadas

### **Monto desembolsado categorizado por zona geográfica.**

La tabla 18 presenta las pruebas de independencia para las variables ZONA GEOGRÁFICA y MONTO DESEMBOLSADO.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables ZONA GEOGRÁFICA y MODE están relacionadas

**Tabla 18** Pruebas de independencia para las variables ZONA GEOGRÁFICA y MODE.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	66,116 <sup>a</sup>	9	0,000

Razón de verosimilitudes	67,036	9	0,000
N de casos válidos	1804		
Nota: a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 25,44			
Fuente los autores, 2017.			

### **Monto desembolsado categorizado por rango de edad.**

La tabla 19 presenta las pruebas de independencia para las variables EDAD y MONTO DESEMBOLSADO.

**Tabla 19** Pruebas de independencia para las variables EDAD y MODE.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	9,330 <sup>a</sup>	6	,156
Razón de verosimilitudes	8,794	6	,185
N de casos válidos	1804		
Nota: a. 1 casillas (8,3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 4,41.			
Fuente los autores, 2017.			

Dado que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy alta, se decide validar la hipótesis de independencia y concluir que las variables EDAD y MODE **NO** están relacionadas.

### **Monto desembolsado categorizado por sexo.**

La tabla 20 presenta las pruebas de independencia para las variables SEXO y MONTO DESEMBOLSADO.

**Tabla 20** Pruebas de independencia para las variables SEXO y MODE.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	78,071 <sup>a</sup>	3	0,000
Razón de verosimilitudes	78,793	3	0,000
N de casos válidos	1804		

---

Nota: a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 47,55.

---

Fuente los autores, 2017.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables SEXO y MODE están relacionadas.

### **Monto desembolsado categorizado por estrato socioeconómico.**

La tabla 21 presenta las pruebas de independencia para las variables ESEC y MONTO DESEMBOLSADO.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables ESEC y MODE están relacionadas.

**Tabla 21** Pruebas de independencia para las variables ESEC y MODE

	<b>Valor</b>	<b>gl</b>	<b>Sig. asintótica (bilateral)</b>
Chi-cuadrado de Pearson	213,552 <sup>a</sup>	12	0,000
Razón de verosimilitudes	155,020	12	0,000
Asociación lineal por lineal	134,946	1	0,000
N de casos válidos	1804		
<i>Nota:</i> a. 5 casillas (25,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es ,66.			

Fuente los autores, 2017.

### **Variables explicativas**

Se realizó análisis de las variables explicativas mediante tablas de contingencia, diagramas de barras y el estadístico chi-cuadrado de Pearson para determinar si están relacionadas.

### **Sexovsestrato socioeconómico.**

La tabla 22 presenta las pruebas de independencia para las variables SEXO yESEC

**Tabla 22** Pruebas de independencia para las variables ESEC y SEXO.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	9,551 <sup>a</sup>	4	0,049
Razón de verosimilitudes	9,532	4	0,049
N de casos válidos	1804		
Nota: a. 1 casillas (10,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 4,80.			

Fuente los autores, 2017.

Se acepta la hipótesis de independencia por tanto las variables SEXO y ESEC NO están relacionadas.

#### **Sexo vs Sector económico.**

La tabla 23 presenta las pruebas de independencia para las variables SEXO y SEEC

**Tabla 23** Pruebas de independencia para las variables SEEC y SEXO.

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	1,224 <sup>a</sup>	3	,747
Razón de verosimilitudes	1,233	3	,745
N de casos válidos	1804		
Nota. a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 6,11			

. Fuente los autores, 2017.

Se acepta la hipótesis de independencia por tanto las variables SEXO y SECTOR ECONÓMICO NO están relacionadas.

#### **Sexo vs Número de trabajadores.**

La tabla 23 presenta las pruebas de independencia para las variables SEXO y NUTR

**Tabla 24** Pruebas de independencia para las variables SEXO y NUTR .

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	52,930 <sup>a</sup>	4	0,000
Razón de verosimilitudes	53,020	4	0,000

N de casos válidos	1804
Nota: a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 75,47.	

Fuente los autores, 2017.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables NUTR y SEXO están relacionadas.

### **Tamaño de muestra**

Se determina el tamaño de muestra de acuerdo a las características de cada diseño muestral.

#### **Tamaño de muestra para M.A.S.**

Se calcula a partir de la formula estándar para tamaño muestral de poblaciones finitas. Los Parámetros del tamaño muestral son: P Confiabilidad 95%, e margen de error del 3%, q complemento para la probabilidad total, Z percentil 95% de la distribución normal a 2 colas, (1(toda la probabilidad) – 2.5=0.975).

- $p < -0.95$
- $q < -1 - p$
- $e < -0.03$
- $z < -qnorm(0.975)$
- Se aplica la formular para calcular tamaño de muestra:
- $n.mas = (N * (z^2) * p * q) / (((e^2) * (N - 1)) + ((z^2) * p * q))$
- $n.mas = 182,3511$
- Se seleccionará una muestra de tamaño 182

#### **Tamaño de muestra para muestreo estratificado.**

Se realizará muestreo estratificado proporcionado con las siguientes como variables de estratificación: SEXO y SECTOR ECONÓMICO. Además dado que no existe información de estudios anteriores sobre el porcentaje de beneficiarios satisfechos con el programa se supondrá  $p=0,5$ ; valor que genera el tamaño de muestra máximo para este diseño muestral.

#### ***Estratificación por sexo.***



La tabla 25 presenta el tamaño de muestra para cada uno de los estratos de la variable SEXO.

**Tabla 25** Tamaño de muestra para muestreo estratificado por SEXO

SEXO	Ni	Pi	Qi	$\frac{Ni*Pi*}{Qi}$	Wi	ni
Hombre	794	0,5	0,5	199	0,44	129
Mujer	1010	0,5	0,5	212	0,56	164
Total	1804			411		293

Fuente los autores, 2017.

Estratificando por SEXO (dos estratos) con un nivel de confianza del 95% y una precisión del 5%, además considerando la información a priori de que el 50% de los hombres y el 50% de las mujeres manifestaron estar satisfechos con las condiciones del programa; se obtiene un tamaño de muestra  $n=293$  dividido en dos estratos  $n_1=129$  (Hombres)  $n_2=164$  (Mujeres).

#### *Estratificación por sector económico.*

La tabla 26 presenta el tamaño de muestra para cada uno de los estratos de la variable SECTOR ECONÓMICO.

**Tabla 26** Tamaño de muestra para muestreo estratificado por SECTOR ECONÓMICO

Sector económico	Ni	Pi	Qi	$Ni*Pi*Qi$	Wi	Ni
Industria*	758	0,5	0,5	190	0,42	126
Servicios	397	0,5	0,5	74	0,22	66
Comercio	649	0,5	0,5	156	0,36	108
Total	1804			420		300

Fuente los autores, 2017.

Estratificando por SECTOR ECONÓMICO (tres estratos) con un nivel de confianza del 95% y una precisión del 5%, además considerando la información a priori de que el 50% del sector industria, 50% del sector servicios y el 50% del sector comercio manifestaron estar satisfechos con las condiciones del programa; se obtiene un tamaño de muestra  $n=300$  dividido en tres estratos  $n_1=126$  (Industria),  $n_2=66$  (Servicios).y  $n_3=108$  (Comercio)

### **Tamaño de muestra para muestreo bietápico por conglomerados y estratificado.**

Se realizarán dos diseños muestrales complejos, en la primera etapa se realizarán conglomerados por zonas geográficas y en la segunda se estratificará por sector socioeconómico en un primer diseño y por sexo en el segundo.

Para la primera etapa de selección de los conglomerados se utilizará una proporción de 1 para garantizar que estén presentes en la muestra unidades de las cuatro zonas geográficas. En la segunda etapa de selección de la muestra por estratos se utilizarán los tamaños de muestra calculados para el muestreo estratificado, los resultados se presentan en las tablas 27 y 28.

**Tabla 27** Tamaño de muestra para segunda etapa de muestreo bietápico estratificado por SEXO.

<b>Estratificación por sexo</b>	<b>Zona Geográfica</b>				
	<b>Centro</b>	<b>Norte</b>	<b>Occidente</b>	<b>Sur</b>	<b>Total</b>
Hombre	31	31	30	37	129
Mujer	39	39	38	47	164
Total	70	70	68	84	293

Fuente los autores, 2017.

**Tabla 28** Tamaño de muestra para segunda etapa de muestreo bietápico estratificado por SECTOR ECONÓMICO.

<b>Estratificación por Sector económico</b>	<b>Zona Geográfica</b>				
	<b>Centro</b>	<b>Norte</b>	<b>Occidente</b>	<b>Sur</b>	<b>Total</b>
Comercio	26	26	25	31	108
Industria	30	30	29	36	126
Servicios	16	16	15	19	66
Total	72	72	70	86	300

Fuente los autores, 2017.

## Planes muestrales y estadísticos

### Muestreo aleatorio simple (M.A.S.).

Se selecciona una muestra de tamaño 182 mediante muestreo aleatorio simple. Las tablas 29 y 30 presentan las unidades muestreadas y las medidas de error para cada una de las categorías de la variable MONTO DESEMBOLSADO.

**Tabla 29** Resumen para muestreo aleatorio M.A.S.

Número de unidades muestreadas		Proporción de unidades muestreadas	
Solicitados	Reales	Solicitados	Reales
182	182	10,1%	10,1%

Fuente los autores, 2017.

**Tabla 30** Medidas del error de muestreo para M.A.S. por categorías de MONTO DESEMBOLSADO.

	Estimación	Error típico	Coefficiente de variación	Efecto del diseño
<= 5 M	822,703	63,327	0,077	1,000
5 M a 11 M	584,813	59,513	0,102	1,000
11 M a 17 M	267,626	45,194	0,169	1,000
17 M +	128,857	32,745	0,254	1,000
Total	1804,000	0,000	0,000	

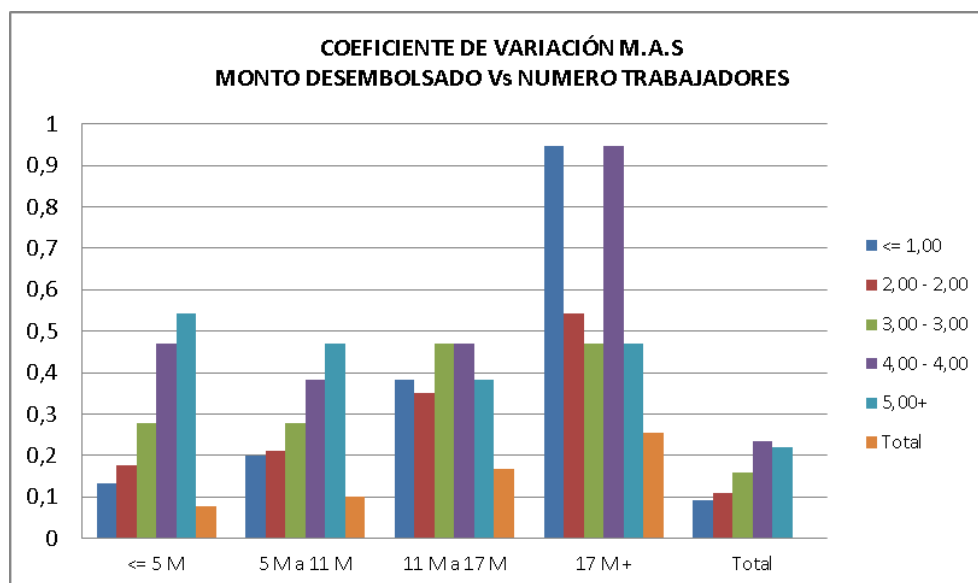
Fuente los autores, 2017.

Los valores obtenidos para los coeficientes de variación dentro de las categorías de MONTO DESEMBOLSADO muestran precisión alta de las estimaciones.

Se elaboran tablas de contingencia entre las variables explicativas y MONTO DESEMBOLSADO, a partir de la muestra para determinar errores de estimación totales y por categoría, así como intervalo de confianza, coeficiente de variación y error de diseño; estadísticos que permitirán comparar la eficiencia de los diseños muestrales. (Ver Apéndice B)

Los siguientes gráficos de barras muestran los coeficientes de variación al interior de las categorías de las variables explicativas NUTR , SEXO y SECTOR ECONÓMICO.

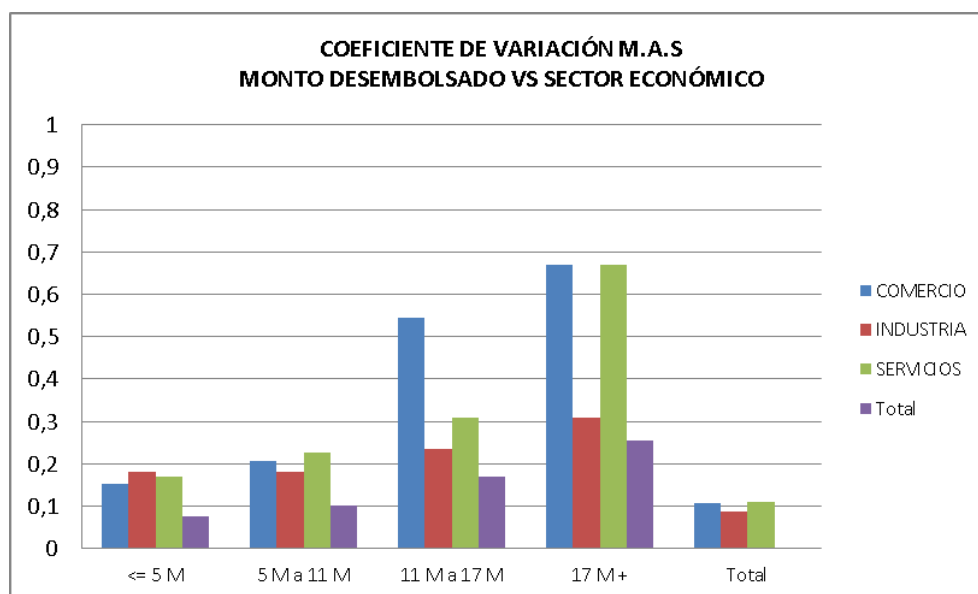
**Figura 2** Gráfico de barras para coeficientes de variación del M.A.S., variables MODE y NUTR.



Fuente los autores, 2017.

Aunque para el total de la muestra el M.A.S se obtienen coeficiente de variación inferiores al 25%, al interior de las categorías de NUMERO TRABAJADORES se observan valores muy altos, en algunos casos superiores al 90%.

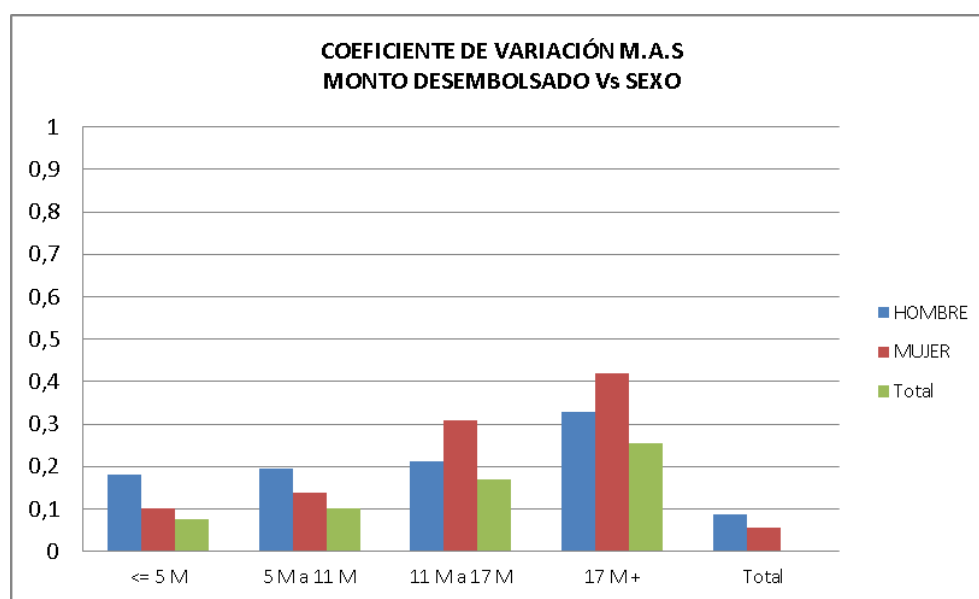
**Figura 3** Gráfico de barras para coeficientes de variación del M.A.S., variables MODE y SEEC.



Fuente los autores, 2017.

Para el total de la muestra el M.A.S se obtienen coeficiente de variación cercanos al 10%, al interior de las categorías de SECTOR ECONÓMICO se observan valores muy altos, en algunos casos superiores al 60%.

**Figura 4** Gráfico de barras para coeficientes de variación del M.A.S., variables MODE y SEXO.



Fuente los autores, 2017.

El M.A.S, para el total de la muestra presenta coeficientes de variación inferiores al 10%, pero al interior de las categorías de SEXO los valores oscilan entre el 20% y el 45%.

### Muestreo estratificado.

Se estratificará por las variables SEXO y SECTOR ECONÓMICO, la variable NUMERO DE TRABAJADORES está relacionada con la variable SEXO por lo que solo se estratificará por SEXO. La variable ESTRATO SOCIOECONÓMICO no es adecuada para estratificar dado que no garantiza que se seleccione al menos una unidad de muestreo de cada estrato

### Muestreo estratificado por sexo (M.E.S.).

Se selecciona una muestra por muestreo estratificado proporcionado, utilizando SEXO como variable de estratificación y tamaño total de muestra 293. Las tablas 301 y 32 presentan los resultados de la selección de la muestra y las medidas de error del diseño muestral aplicado.

**Tabla 31** Resumen para muestreo estratificado por SEXO (M.E.S.).

		<b>Número de unidades muestreadas</b>		<b>Proporción de unidades muestreadas</b>	
		<b>Solicitados</b>	<b>Reales</b>	<b>Solicitados</b>	<b>Reales</b>
Sexo_ =	Hombre	129	129	16,4%	16,4%
	Mujer	164	164	16,1%	16,1%

Fuente los autores, 2017.

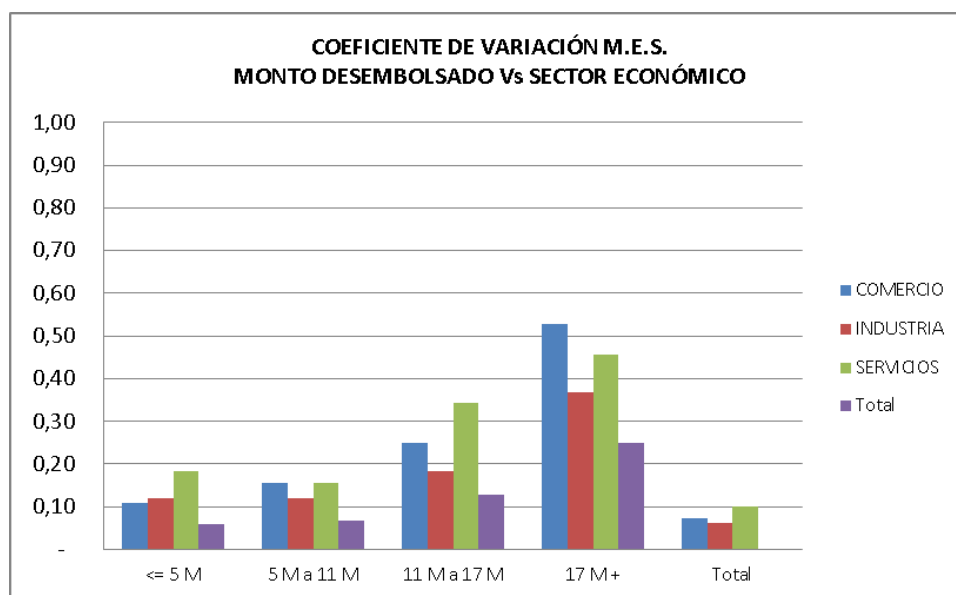
**Tabla 32** Medidas del error de muestreo para M.E.S. categorizado por MONTO DESEMBOLSADO.

<b>MONTO DESEMBOLSADO</b>	<b>Estimación</b>	<b>Error típico</b>	<b>Coefficiente de variación</b>	<b>Efecto del diseño</b>
<= 5 M	783,64	46,66	0,06	0,95
5 M a 11 M	676,41	46,50	0,07	0,99
11 M a 17 M	264,14	33,94	0,13	0,99
17 M +	79,81	19,82	0,25	1,00
Total	1.804,00	-	-	.

Fuente los autores, 2017.

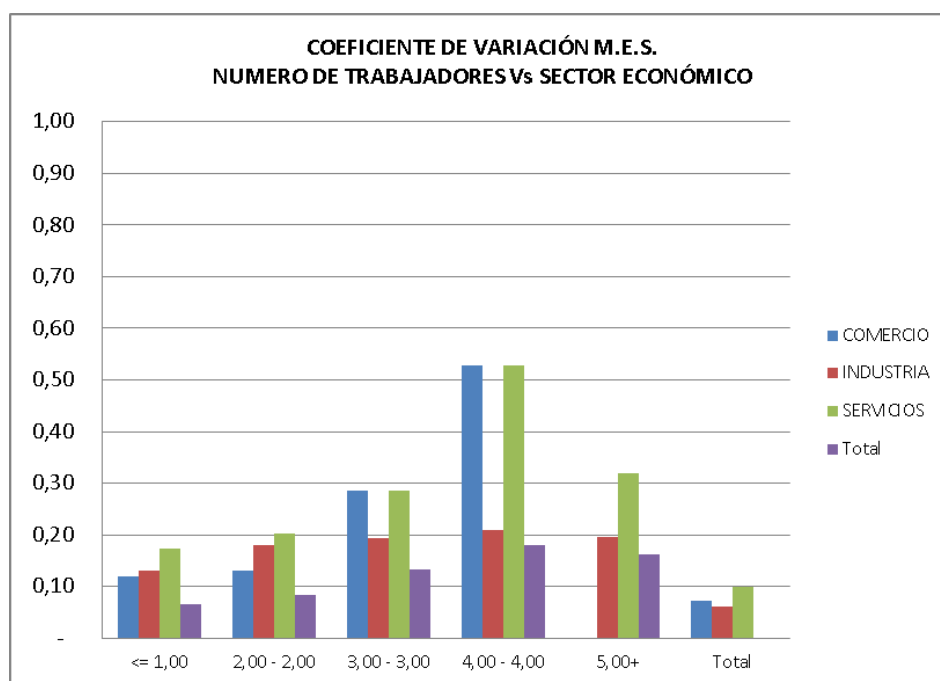
Se elaboran tablas de contingencia a partir de la muestra para determinar errores de por categoría, así como coeficiente de variación y error de diseño; estadísticos que permitirán comparar la eficiencia de los diseños muestrales (Ver Apéndice C). Las figuras 5 y 6 muestran en gráficos de barras la variabilidad al interior de las clases de las variables MONTO DESEMBOLSADO y NUMERO TRABAJADORES.

**Figura 5** Gráfico de barras para coeficientes de variación del M.E.S., variables MODE y SEEC.



Fuente los autores, 2017.

**Figura 6** Gráfico de barras para coeficientes de variación del M.E.S., variables NUTR y SEEC.



Fuente los autores, 2017.

Revisando el error típico y el coeficiente de variación se observa que para el total de casos el muestreo estratificado por SEXO genera muestras bastante precisas, pero al revisar al interior de cada una de las categorías de SECTOR ECONÓMICO se observa que para montos

desembolsados superiores a 11 millones la variabilidad es muy alta. El efecto de diseño se tendrá en cuenta al momento de comparar los diseños muestrales.

La estimación del número de trabajadores resulto un poco más precisa para el total de los casos, pero al interior de las categorías de SECTOR ECONÓMICO continúa presentándose alta variabilidad.

***Muestreo estratificado por sector económico (M.E.S.E.).***

Se selecciona una muestra por muestreo estratificado proporcionado, utilizando SECTOR ECONÓMICO como variable de estratificación y tamaño total de muestra. Las tablas 33 y 34 presentan los resultados y las medidas de error para el M.E.S.E.

**Tabla 33** Resumen para muestreo estratificado por SECTOR ECONÓMICO.

Sector económico	Número de unidades muestradas		Proporción de unidades muestradas	
	Solicitados	Reales	Solicitados	Reales
Comercio	108	108	16,8%	16,8%
Industria	126	126	16,7%	16,7%
Servicios	66	66	16,3%	16,3%

Fuente los autores, 2017.

**Tabla 34** Medidas del error de muestreo para M.E.S.E. categorizado por MONTO DESEMBOLSADO.

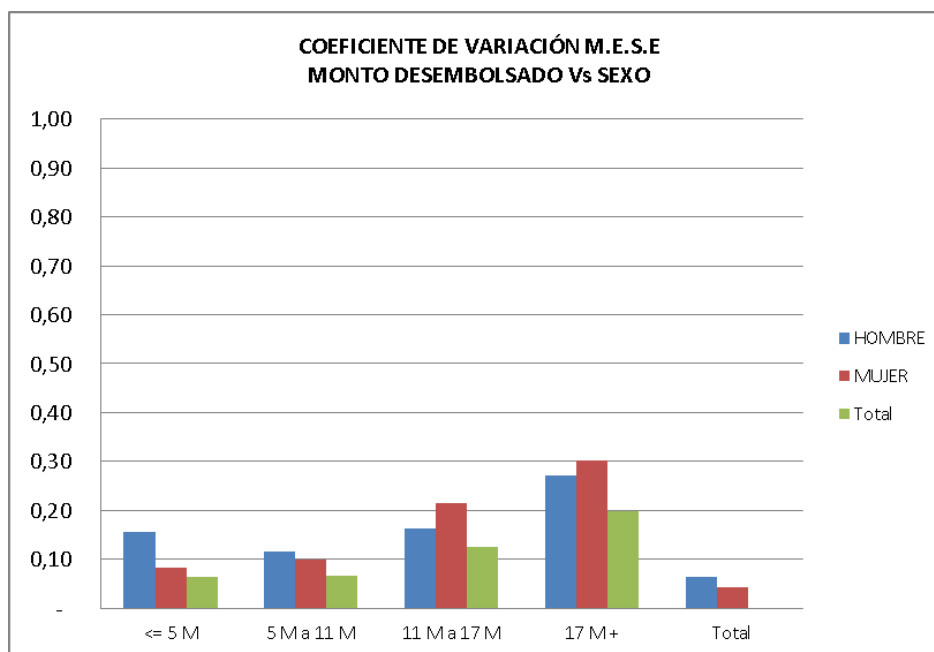
	Estimación	Error típico	Coefficiente de variación	Efecto del diseño
<= 5 M				
	708,21	46,16	0,07	0,98
5 M a 11 M				
	703,98	46,50	0,07	1,00
11 M a 17 M				
	271,29	33,78	0,12	0,98
17 M +				
	120,53	23,87	0,20	1,01
Total				
	1.804,00	-	-	-

Fuente los autores, 2017.

Las figuras 7 y 8 muestran los gráficos de barras de los coeficientes de variación al interior de las clases de las variables MODE, SEEC y SEXO.

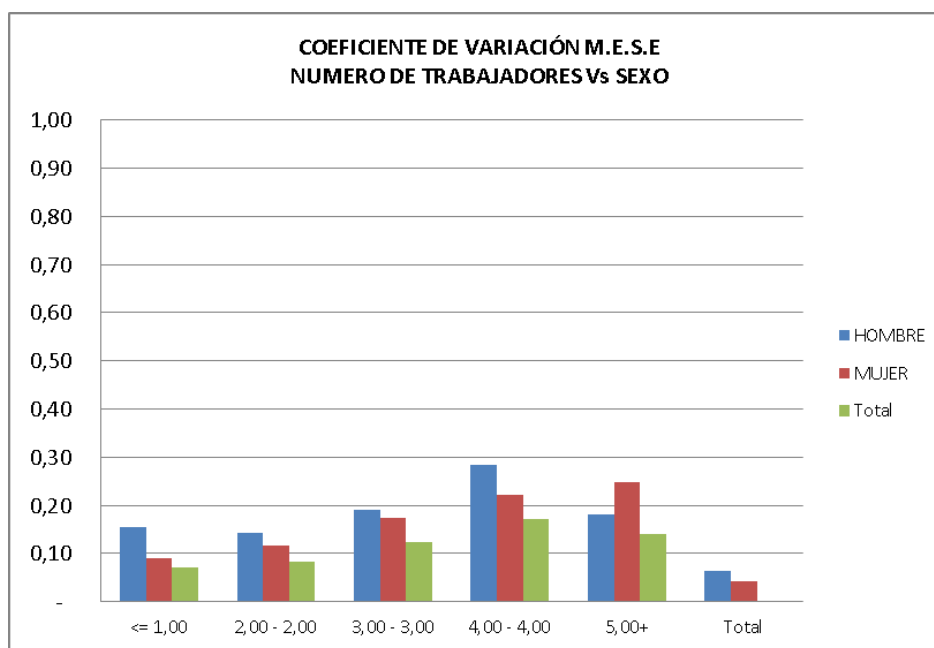


**Figura 7** Gráfico de barras para coeficientes de variación del M.E.S.E., variables MODE y SEXO.



Fuente los autores, 2017.

**Figura 8** Gráfico de barras para coeficientes de variación del M.E.S.E., variables NUTR y SEXO



Fuente los autores, 2017.

Revisando el error típico y el coeficiente de variación se observa que para el total de casos el muestreo estratificado por SECTOR ECONÓMICO genera muestras precisas, en todas las clases, al revisar al interior de cada una de las categorías de SEXO el Coeficiente de variación (CV) es inferior al 30%

La estimación del número de trabajadores resultó aún más precisa para el total de los casos, y al interior de las categorías de SEXO también disminuye la variabilidad.

### ***Muestreo complejo por conglomerados y estratificado.***

Se realizaron dos diseños muestrales bietápicos de la siguiente forma: para el primer diseño se tomaron conglomerados en la primera etapa usando la variable ZONA GEOGRÁFICA y se estratificó en la segunda etapa por la variable SEXO; para el segundo diseño se tomaron conglomerados en la primera etapa usando la variable ZONA GEOGRÁFICA y se estratificó en la segunda etapa por la variable SECTOR ECONÓMICO.

### ***Muestra 01: Conglomerada por zona geográfica y estratificada por sector económico.***

Se seleccionó una muestra de 299 unidades, para cada conglomerado y estrato se tomó una submuestra proporcional al tamaño que presentan en la población.

Las tablas 35 y 36 presentan los resultados de las etapas 1 y 2 para la muestra 01.

**Tabla 35** Resumen para la etapa 1 conglomerados por ZONA GEOGRÁFICA.

Número de unidades muestradas		Proporción de unidades muestradas	
Solicitados	Reales	Solicitados	Reales
4	4	100,0%	100,0%

Fuente los autores, 2017.

**Tabla 36** Resumen para la etapa 2 estratos por SECTOR ECONÓMICO.

Zona geográfica	Sector económico	Número de unidades muestradas		Proporción de unidades muestradas	
		Solicitados	Reales	Solicitados	Reales
Centro	Comercio	20	20	16,6%	16,8%
	Industria	37	37	16,6%	16,8%
	Servicios	15	15	16,6%	16,5%
Norte	Comercio	25	25	16,6%	16,7%
	Industria	28	28	16,6%	16,8%
	Servicios	19	19	16,6%	16,4%
Occidente	Comercio	35	35	16,6%	16,6%

	Industria	19	19	16,6%	16,2%
	Servicios	15	15	16,6%	16,1%
	Comercio	27	27	16,6%	16,5%
Sur	Industria	42	42	16,6%	16,7%
	Servicios	17	17	16,6%	16,2%

Fuente los autores, 2017.

Se seleccionó una muestra de tamaño 299 distribuida en cuatro zonas geográficas y tres estratos por zona. El tamaño de muestra de cada estrato por zona es proporcional al tamaño que ocupa en la población.

La tabla 37 presenta las medidas de error para el muestreo bietápico estratificado por SECTOR ECONÓMICO.

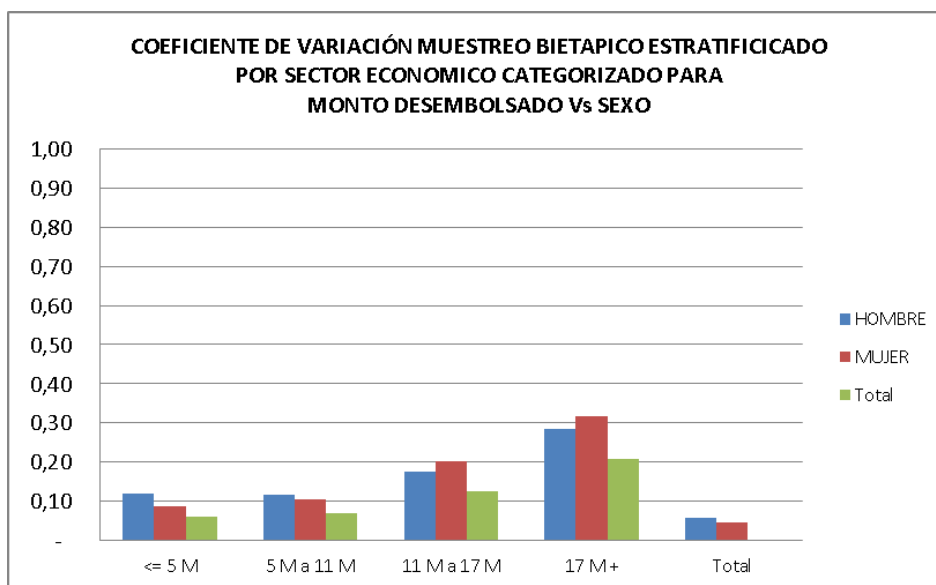
**Tabla 37** Medidas de error para muestreo bietápico estratificado por SECTOR ECONÓMICO.

<b>MONTO DESEMBOLSADO</b>	<b>Estimación</b>	<b>Error típico</b>	<b>Coefficiente de variación</b>	<b>Efecto del diseño</b>
<= 5 M				
	767,36	46,13	0,06	0,96
5 M a 11 M				
	664,12	45,63	0,07	0,98
11 M a 17 M				
	264,52	33,30	0,13	0,97
17 M +				
	107,99	22,41	0,21	0,98
Total				
	1.804,00	-	-	-

Fuente los autores, 2017.

La figura 9 muestra el gráfico de barras de los coeficientes de variación al interior de las clases de la variable SEXO.

**Figura 9** Gráfico de barras para coeficientes de variación del muestreo bietápico estratificado por SECTOR ECONÓMICO.



Fuente los autores, 2017.

Se observa en el gráfico anterior que la variabilidad es mínima para el total de la muestra (coeficiente de variación menor al 6%), pero para la categoría 17 M o más de la variable MONTO DESEMBOLSADO se presentan coeficientes de variación cercanos al 30%.

***Muestra 02: conglomerada por zona geográfica y estratificada por sexo.***

Se seleccionó una muestra de 291 unidades, para cada conglomerado y estrato se tomó una submuestra proporcional al tamaño que presentan en la población.

Las tablas 38 y 39 presentan los resultados de las etapas 1 y 2 para la muestra 02.

**Tabla 38** la etapa 1 conglomerados por ZONA GEOGRÁFICA.

Número de unidades		Proporción de unidades	
<i>Resumen para muestreadas</i>		<i>muestreadas</i>	
Solicitados	Reales	Solicitados	Reales
4	4	100,0%	100,0%

Fuente los autores, 2017.

**Tabla 39** Resumen para la etapa 2 estratificada por SEXO.

Zona geográfica	Sexo	Número de unidades muestreadas		Proporción de unidades muestreadas	
		Solicitados	Reales	Solicitados	Reales
Centro	Hombre	35	35	16,2%	16,1%

Norte	Mujer	34	34	16,2%	16,0%
	Hombre	31	31	16,2%	16,2%
	Mujer	39	39	16,2%	16,1%
Occidente	Hombre	26	26	16,2%	16,4%
	Mujer	42	42	16,2%	16,0%
Sur	Hombre	35	35	16,2%	16,0%
	Mujer	49	49	16,2%	16,3%

Fuente los autores, 2017.

Se seleccionó una muestra de 291 unidades, para cada conglomerado y estrato se tomó una submuestra proporcional al tamaño que presentan en la población.

La tabla 40 presenta las medidas de error para el muestreo bietápico estratificado por SEXO.

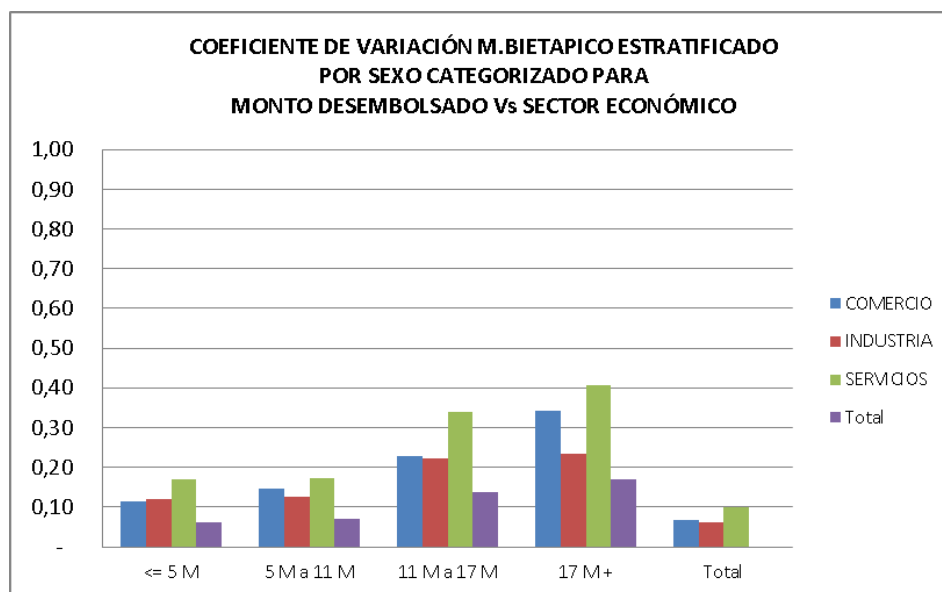
**Tabla 40** Medidas de error para muestreo bietápico estratificado por SEXO.

Monto desembolsado	Estimación	Error típico	Coefficiente de variación	Efecto del diseño
<= 5 M	756,69	46,10	0,06	0,93
5 M a 11 M	650,59	46,33	0,07	0,99
11 M a 17 M	235,44	32,66	0,14	1,00
17 M +	161,28	27,29	0,17	0,97
Total	1.804,00	0,00	0,00	

Fuente los autores, 2017.

La figura 10 muestra el gráfico de barras de los coeficientes de variación al interior de las clases de la variable SECTOR ECONÓMICO.

**Figura 10** Gráfico de barras para coeficientes de variación del muestreo bietápico estratificado por SEXO.



Variables MODE y SECTOR ECONÓMICO. Fuente los autores, 2017.

Al interior de las clases de MONTO DESEMBOLSADO se obtienen coeficientes de variación entre el 5% y el 25% con tres excepciones que superan el 30%.

### Discusión de diseños muestrales

Para determinar cuál de los diseños permite seleccionar la mejor muestra (más representativa) comparamos la variabilidad (Coeficiente de variación) para las diferentes categorías de la variable MONTO DESEMBOLSADO y al interior de las clases de las variables explicativas SEXO, SECTOR ECONÓMICO y NUMERO DE TRABAJADORES.

Tabla 41 Categorías de las variables explicativas

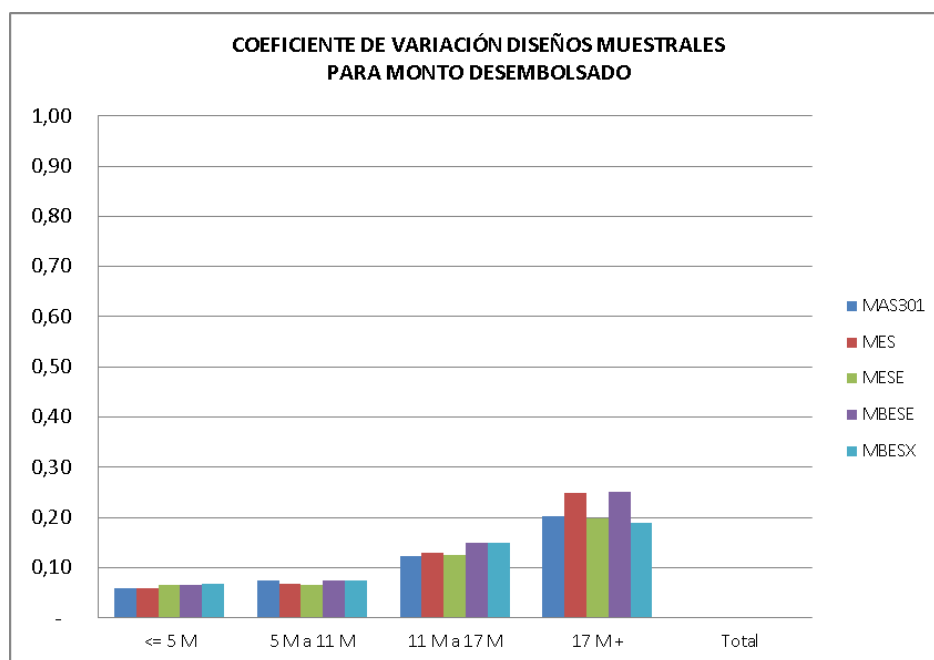
Variable	Tipo de variable	Categorías
MONTO DESEMBOLSADO	Dependiente	Menor a 5 millones 5 a 11 millones 11 a 17 millones 17 millones o más
SEXO	Explicativa	Hombre Mujer
SECTOR ECONOMICO	Explicativa	Comercio, Industria, Servicios
NUMERO DE TRABAJADORES	Explicativa	

Fuente los autores, 2017

La figura11 presenta los coeficientes de variación de los diferentes diseños muestrales al interior de las categorías de MONTO DESEMBOLSADO. Las siglas de la leyenda del gráfico corresponden a los siguientes diseños muestrales:

- MAS301: Muestreo aleatorio simple con tamaño de muestra  $n=301$
- MES: Muestreo estratificado por SEXO
- MESE: Muestreo estratificado por SECTOR ECONOMICO
- MBSE: Muestreo bietápico estratificado por SECTOR ECONOMICO
- MBSX: Muestreo bietápico estratificado por SEXO

**Figura 11** Gráfico de barras para coeficientes de variación de los diseños muestrales.



Fuente los autores, 2017.

Revisando la variabilidad únicamente para las clases de MONTO DESEMBOLSADO, sin considerar el comportamiento al interior de las categorías de las variables explicativas se observa que se obtienen valores muy cercanos del coeficiente de variación, pero la menor variabilidad para la categoría de 17 millones o más se presenta para el diseño bietápico estratificado por SEXO.

Dado que es de interés para el programa de inversión, por su carácter social, tener representación de las diferentes subpoblaciones al interior de la muestra; el coeficiente de variación para las clases de MONTO DESEMBOLSADO no es criterio suficiente para seleccionar el mejor diseño muestral.

Se realizó análisis de las variables explicativas mediante tablas de contingencia, diagramas de barras y el estadístico chi-cuadrado de Pearson determinando que las variables SEXO y SECTOR ECONOMICO no están correlacionadas entre sí pero están altamente correlacionadas con la variable MONTO DESEMBOLSADO por lo cual se consideran los coeficientes de variación al interior de las clases de las variables SEXO y SECTOR ECONÓMICO como criterios adicionales de selección del mejor diseño muestral, (Figura 9 y Figura 10). Siendo estos los criterios el diseño que selecciona la muestra más representativa de la población es el Muestreo Bietapico conglomerado por ZONA GEOGRÁFICA y estratificado por la variable SEXO, las medidas de error para este plan muestral se presentan en la Tabla 40



## Conclusiones de la investigación

El 65% de los beneficiados con créditos para el fortalecimiento de los proyectos productivos son personas naturales, adultos, heterosexuales, no son desplazados ni pertenecen a algún grupo étnico y tienen como máximo estudios secundarios completos, es decir el proyecto de financiamiento no está cumpliendo el objetivo de beneficiar a población especial o vulnerable.

Para determinar el mejor diseño muestral no es suficiente comparar los coeficientes de variación obtenidos para la variable MONTO DESEMBOLSADO ya que todos los diseños arrojan valores muy cercanos al interior de las categorías y de cero para el total de la estimación de la variable.

Si el único criterio de selección de un diseño muestral es la variabilidad, generalmente el M.A.S. resulta la mejor opción.

Al utilizar un muestreo estratificado es imprescindible garantizar que el diseño asegure que se seleccione al menos una unidad de muestreo de cada estrato.

Las unidades de muestreo al interior de cada estrato deben ser muy similares para garantizar la homogeneidad.

Al momento de seleccionar el diseño es importante considerar la intención del muestreo, por ejemplo si es para realizar encuestas, los costos de aplicación de las mismas resultan de gran importancia por lo que podría pensarse en seleccionar previamente conglomerados.

El muestreo estratificado reduce la probabilidad de que se seleccione un número desproporcionado de unidades de una subpoblación que se considere importante para el estudio que se esté realizando.

Las unidades de muestreo al interior de cada conglomerado deben ser lo más heterogéneas posible.

La conglomeración de las unidades en una etapa reduce los costos pero aumenta la varianza de las estimaciones debido a las correlaciones que se presentan entre las unidades del mismo conglomerado.

Para el presente estudio el diseño muestral más adecuado resulta ser el bietápico por conglomerados y estratificado por SEXO no solo por presentar muy poca variabilidad (coeficiente de variación) al interior de las clases de la variable MONTO DESEMBOLSADO si no que el hecho de utilizar conglomerados reduce los costos de aplicación de las encuestas y garantiza representatividad de toda la población.

## Referencias

- Abascal, E., & Esteban, I. G. (2005). *Análisis de encuestas*. ESIC Editorial.
- Alvarado Jorge. (2008). *Fundamentos de inferencia estadística*. Pontificia Universidad Javeriana.
- Azorín, F., Poch, F. A., & Sánchez-Crespo, J. L. (1986). *Métodos y aplicaciones del muestreo*. Alianza Editorial.
- Berenson, M. L., & Levine, D. M. (1996). *Estadística básica en administración: conceptos y aplicaciones*. Pearson Educación.
- Camara de Comercio de Bogota. (2012). ¿Cómo financiar tu empresa?-Consejos prácticos.
- Cardenas Santa Maria Maurico. (2013). *Introducción a la economía colombiana*. Alfaomega Colombiana.
- Consejo de Bogotá, C. (2012). Plan de desarrollo económico, social, ambiental y de obras públicas para Bogotá DC 2012-2016. *Bogotá: Oficina de Comunicaciones, Alcaldía Mayor de Bogotá*.
- Freund, J. E., & Simon, G. A. (1994). *Estadística elemental*. Pearson Educación.
- García, M. A. (1990). *Estadística avanzada con el paquete Systat*. EDITUM.
- Hernández, M. A., Cantin Garcia, S., Lopez Abejon, N., & Rodriguez Zazo, M. (2014). Estudio de encuestas. *Recuperado en abril*. Recuperado a partir de [https://www.uam.es/personal\\_pdi/stmaria/jmurillo/InvestigacionEE/Presentaciones/Curso\\_10/ENCUESTA\\_Trabajo.pdf](https://www.uam.es/personal_pdi/stmaria/jmurillo/InvestigacionEE/Presentaciones/Curso_10/ENCUESTA_Trabajo.pdf)
- Kendall, K. E., & Kendall, J. E. (1997). *Análisis y diseño de sistemas*. Pearson Educación.
- Levin, R. I., & Rubin, D. S. (2004). *Estadística para administración y economía*. Pearson Educación.
- Malhotra, N. K. (2004). *Investigación de mercados: un enfoque aplicado*. Pearson Educación.
- Misas Gabriel. (2004). *La educación superior en Colombia: análisis y estrategias para su desarrollo*. Univ. Nacional de Colombia.
- Moore, D. S. (2005). *Estadística aplicada básica*. Antoni Bosch editor.
- Morera Rojas, A. (2013). La incidencia del modelo de economía de mercado en la asignación del gasto público en Bogotá durante los últimos veinte años.
- Niño Riaño, N., & Velandia, D. A. (2011). Estudio del costo beneficio del microcredito dirigido a los microempresarios de Bogotá 2000-2006.
- Pacheco Espejel, A. A. (2013). La estadística aplicada como herramienta para la dirección integral de las organizaciones.
- Peña, D. (2008). *Fundamentos de estadística*. Alianza Editorial.
- Perez Rigoberto. (2010). *Nociones Básicas de Estadística*. Rigoberto Perez.
- Pineda, O. L. (1998). *Métodos y modelos econométricos: una introducción*. Editorial Limusa.
- Ruiz, C. Q. (1993). *Elementos de Inferencia Estadística*. Editorial Universidad de Costa Rica.
- Sabadías, A. V. (1995). *Estadística descriptiva e inferencial*. Univ de Castilla La Mancha.
- Salkind, N. J. (1999). *Métodos de investigación*. Pearson Educación.
- Sanchez Perez, G., Duarte, L., & Blanco, M. (2013). El conocimiento en la economía global y colombiana. *Revista Facultad de Ciencias Económicas: Investigación y Reflexión*, 21(2), 97-114.
- Santofimio Vargas, F. (2016). Propuesta de una herramienta para identificar los determinantes del fracaso en unidades productivas de poblaciones jóvenes en condición de Vulnerabilidad de Bogotá DC.

- Scheaffer, R. L., Mendenhall, W., & Ott, L. (2006). *Elementos de muestreo*. Editorial Paraninfo.
- Taborga, C. E. V., Castellón, R. V., & Taborga, O. Á. V. (2012). DETERMINACIÓN DEL TAMAÑO MUESTRAL MEDIANTE EL USO DE ÁRBOLES DE DECISIÓN SAMPLE SIZE DETERMINATION USING DECISION TREES. *Revista Investigación & Desarrollo*, 1(11). Recuperado a partir de <http://www.upb.edu/revista-investigacion-desarrollo/index.php/id/article/view/64>
- Turriago, C. E. P. (2011). *Métodos en ejes principales para tablas de contingencia con estructuras de partición en filas*. Universidad Nacional de Colombia.
- Vázquez, A. L., & Ortiz, F. J. G. (2005). *Métodos estadísticos para medir, describir y controlar la variabilidad*. Ed. Universidad de Cantabria.
- Vivanco, M. (2005). *Muestreo Estadístico. Diseño Y Aplicaciones*. Editorial Universitaria.
- Zapata-Ossa, H. de J., Cubides-Munévar, A. M., López, M. C., Pinzón-Gómez, E. M., Filigrana-Villegas, P. A., & Cassiani-Miranda, C. A. (2011). Muestreo por conglomerados en encuestas poblacionales. *Revista de Salud Pública*, 13(1), 141–151.

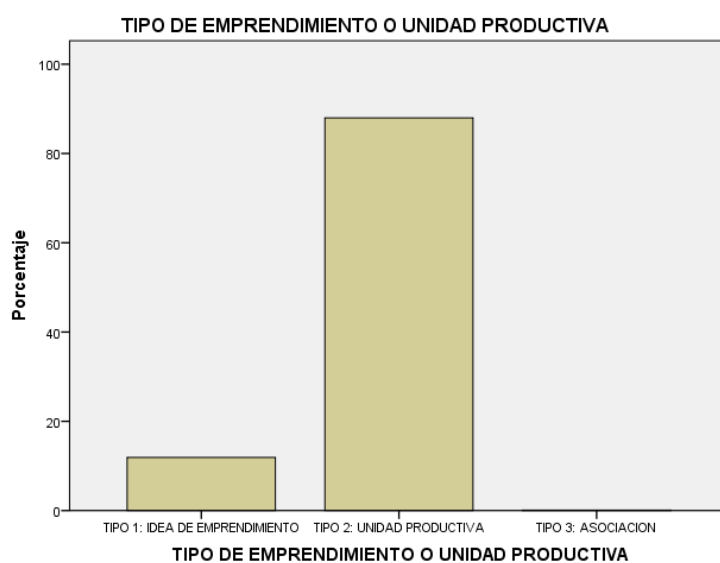
## APÉNDICE A

### ANÁLISIS DESCRIPTIVO

Se construyeron tablas de contingencia y diagramas de barras para analizar el comportamiento de las 9 variables categóricas

#### TIPO DE EMPRENDIMIENTO O UNIDAD PRODUCTIVA:

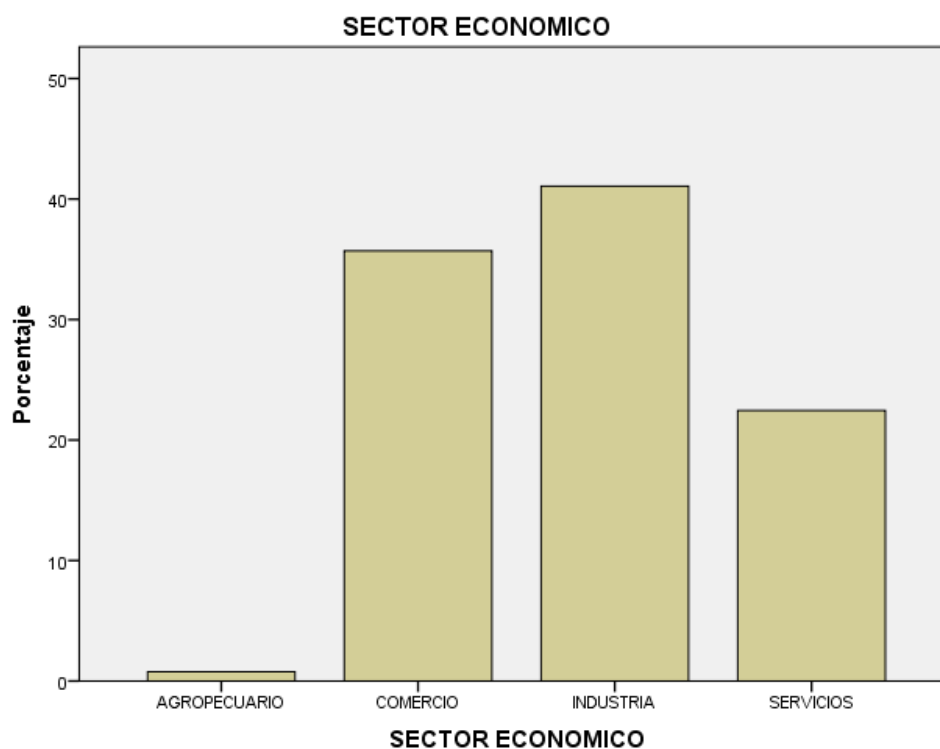
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
TIPO 1: IDEA DE EMPRENDIMIENTO	215	11,9	11,9	11,9
TIPO 2: UNIDAD PRODUCTIVA	1587	88,0	88,0	99,9
TIPO 3: ASOCIACIÓN	2	,1	,1	100,0
Total	1804	100,0	100,0	



El 88% de los beneficiarios de crédito son unidades productivas por tanto no es una variable que represente importancia para el análisis del MONTO DESEMBOLSADO

**SECTOR ECONÓMICO:**

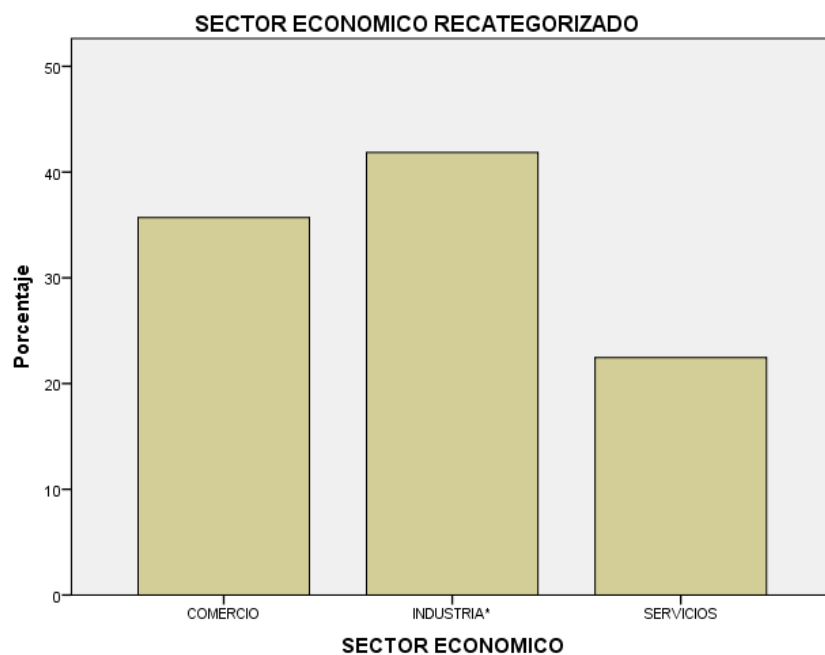
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje Acumulado
Válidos	AGROPECUARIO	14	,8	,8	,8
	COMERCIO	644	35,7	35,7	36,5
	INDUSTRIA	741	41,1	41,1	77,5
	SERVICIOS	405	22,5	22,5	100,0
	Total	1804	100,0	100,0	



Dado que Bogotá es en su mayoría urbana, menos de un 1% de los beneficiados pertenecen al sector agropecuario. Resulta conveniente incluir el sector agropecuario dentro del sector industrial

**SECTOR ECONÓMICO MODIFICADO**

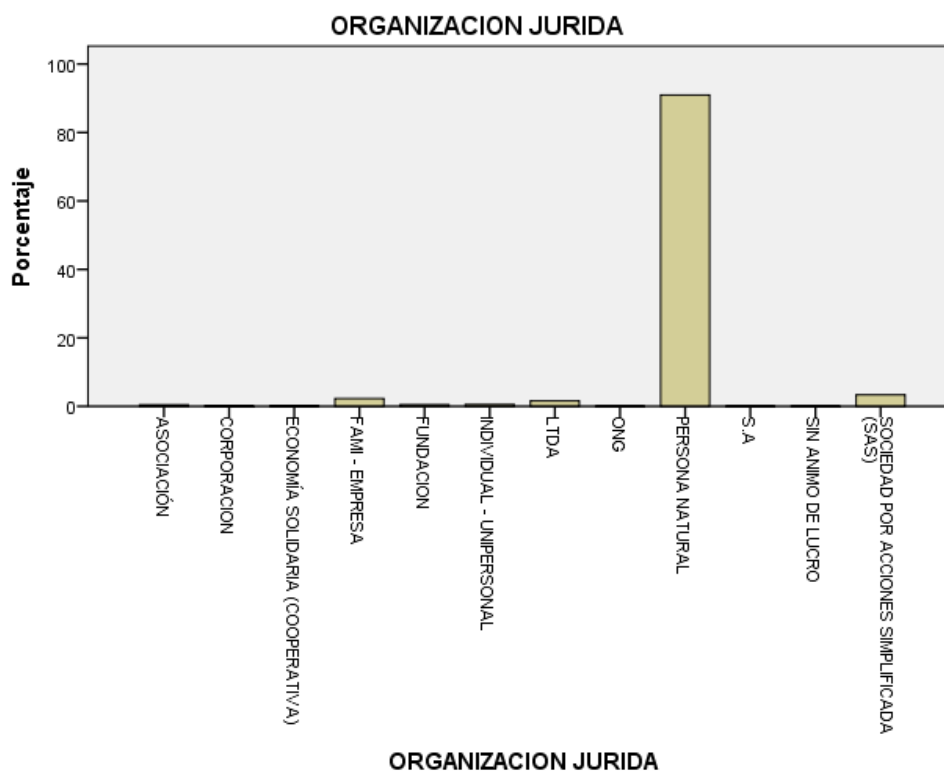
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	COMERCIO	644	35,7	35,7	35,7
	INDUSTRIA	755	41,9	41,9	77,5
	SERVICIOS	405	22,5	22,5	100,0
	Total	1804	100,0	100,0	



\*Se redujo una categoría (Sector Agropecuario) debido al reducido número de casos, el sector INDUSTRIA\* agrupa los sectores INDUSTRIA y AGROPECUARIO

**ORGANIZACIÓN JURÍDICA:**

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje Acumulado
ASOCIACIÓN	8	,4	,4	,4
CORPORACIÓN	2	,1	,1	,6
ECONOMÍA SOLIDARIA (COOPERATIVA)	2	,1	,1	,7
FAMI - EMPRESA	41	2,3	2,3	2,9
FUNDACIÓN	9	,5	,5	3,4
INDIVIDUAL - UNIPERSONAL	10	,6	,6	4,0
VálidosLTDA	28	1,6	1,6	5,5
ONG	1	,1	,1	5,6
PERSONA NATURAL	1641	91,0	91,0	96,6
S.A	1	,1	,1	96,6
SIN ANIMO DE LUCRO	1	,1	,1	96,7
SOCIEDAD POR ACCIONES SIMPLIFICADA (SAS)	60	3,3	3,3	100,0
Total	1804	100,0	100,0	

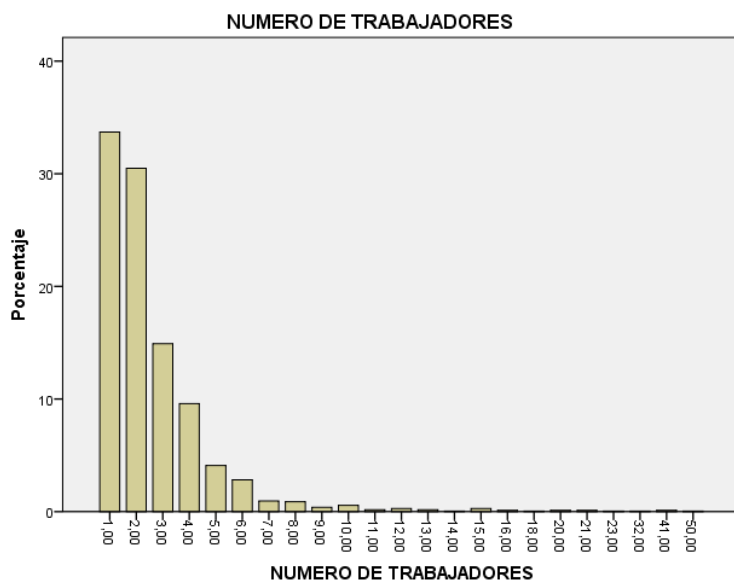


El 91% de los beneficiados con créditos son personas jurídicas, por tanto no es una variable que represente importancia para el análisis del MONTO DESEMBOLSADO

**NÚMERO DE TRABAJADORES:**

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
1,00	608	33,7	33,7	33,7
2,00	550	30,5	30,5	64,2
3,00	269	14,9	14,9	79,1
4,00	173	9,6	9,6	88,7
5,00	74	4,1	4,1	92,8
6,00	51	2,8	2,8	95,6
7,00	17	,9	,9	96,6
8,00	16	,9	,9	97,5
9,00	7	,4	,4	97,8
10,00	10	,6	,6	98,4
11,00	3	,2	,2	98,6
12,00	5	,3	,3	98,8
13,00	3	,2	,2	99,0
14,00	1	,1	,1	99,1
15,00	5	,3	,3	99,3
16,00	2	,1	,1	99,4
18,00	1	,1	,1	99,5
20,00	2	,1	,1	99,6
21,00	2	,1	,1	99,7
23,00	1	,1	,1	99,8
32,00	1	,1	,1	99,8
41,00	2	,1	,1	99,9
50,00	1	,1	,1	100,0
Total	1804	100,0	100,0	

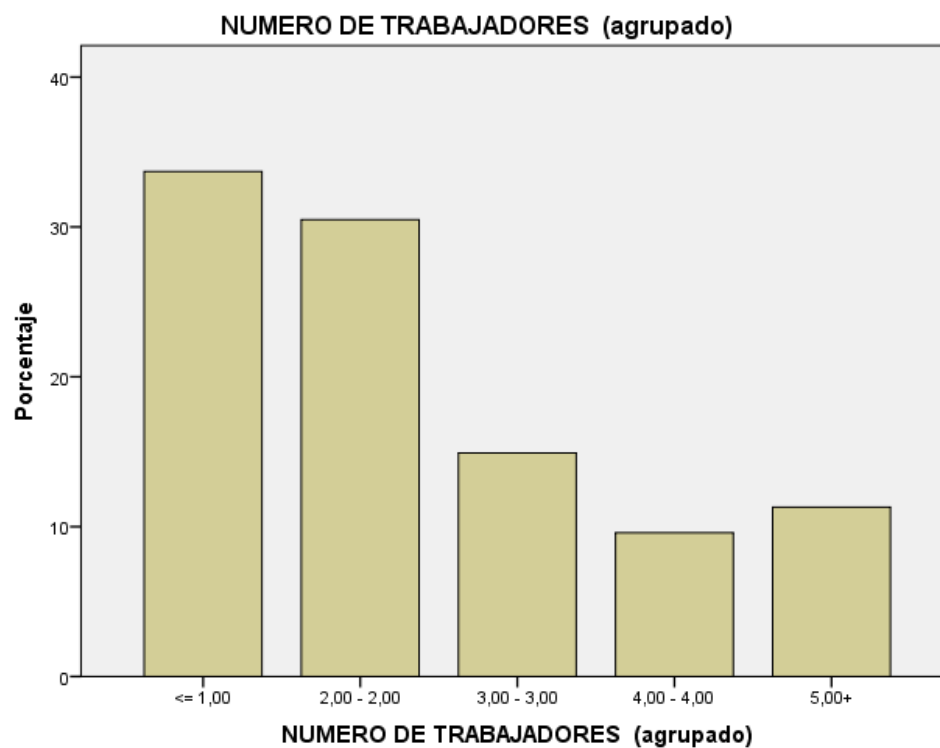




Dado que el 98,4% de las unidades productivas tienen 10 o menos empleados, se hace necesario reagrupar para obtener menor número de clases. Se crean 5 clases de la siguiente forma:

**NUMERO DE TRABAJADORES (agrupado)**

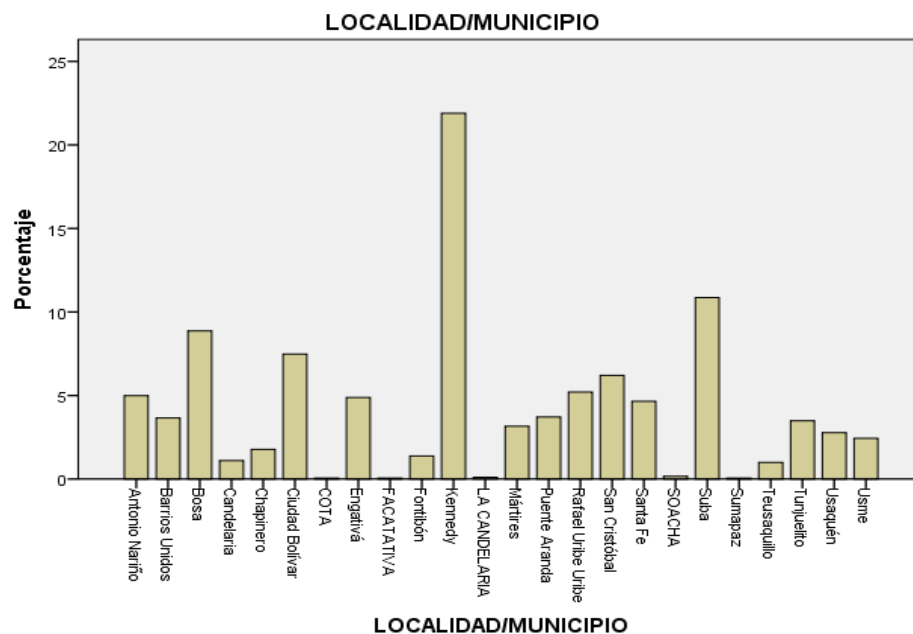
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<= 1,00	608	33,7	33,7	33,7
2,00 - 2,00	550	30,5	30,5	64,2
3,00 - 3,00	269	14,9	14,9	79,1
4,00 - 4,00	173	9,6	9,6	88,7
5,00+	204	11,3	11,3	100,0
Total	1804	100,0	100,0	



Agrupando en 5 clases se observa en el diagrama de barras que más del 60% de las unidades de negocio tienen a lo sumo dos trabajadores y que apenas un 11% tienen cinco o más empleados.

**LOCALIDAD/MUNICIPIO**

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Antonio Nariño	90	5,0	5,0
	Barrios Unidos	66	3,7	8,6
	Bosa	160	8,9	17,5
	Candelaria	20	1,1	18,6
	Chapinero	32	1,8	20,4
	Ciudad Bolívar	135	7,5	27,9
	COTA	1	,1	27,9
	Engativá	88	4,9	32,8
	FACATATIVA	1	,1	32,9
	Fontibón	25	1,4	34,3
	Kennedy	395	21,9	56,2
	LA CANDELARIA	2	,1	56,3
	Mártires	57	3,2	59,4
	Puente Aranda	67	3,7	63,1
	Rafael Uribe Uribe	94	5,2	68,3
	San Cristóbal	112	6,2	74,6
	Santa Fe	84	4,7	79,2
	SOACHA	3	,2	79,4
	Suba	196	10,9	90,2
	Sumapaz	1	,1	90,3
	Teusaquillo	18	1,0	91,3
	Tunjuelito	63	3,5	94,8
	Usaquén	50	2,8	97,6
	Usme	44	2,4	100,0
	Total	1804	100,0	



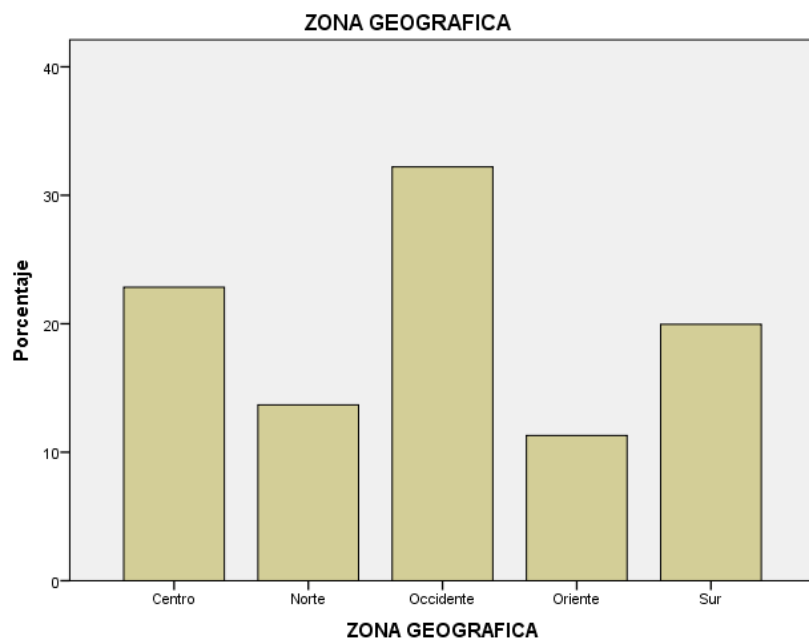
El 21,9% de los beneficiados se encuentran ubicados en la localidad de Kennedy, siendo esta la de mayor representación.

De trabajar con la variable localidad sería necesario agrupar localidades para lo cual es importante realizarlo geográficamente.

Se agrupan las localidades y municipios aledaños en 5 zonas geográficas con las siguientes frecuencias de beneficiarios

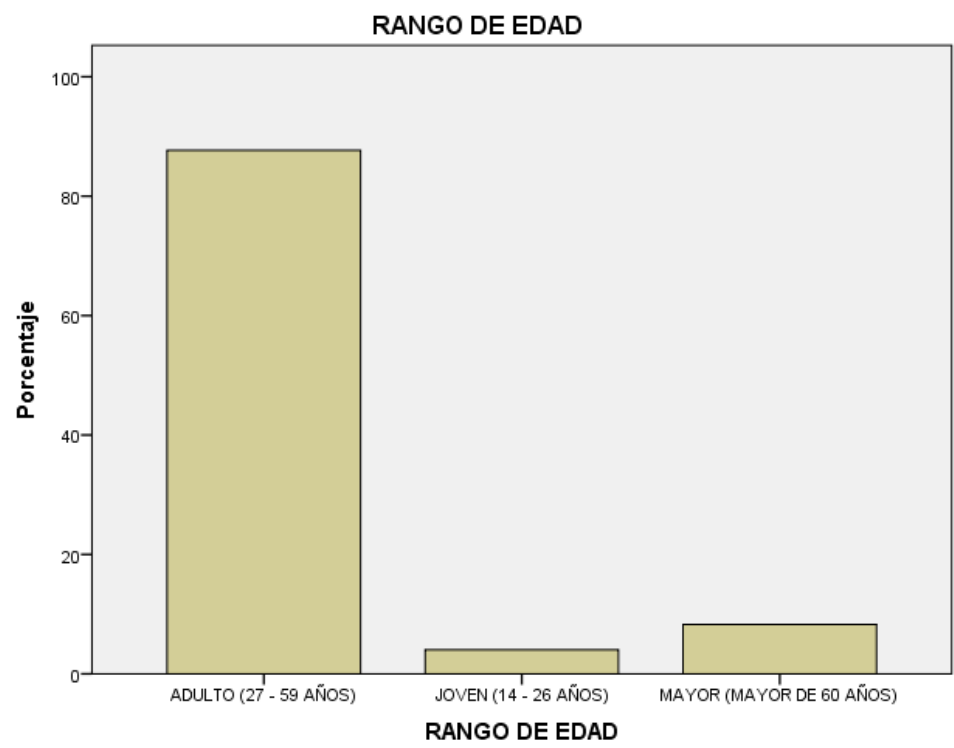
#### ZONA GEOGRAFICA

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Centro	412	22,8	22,8	22,8
	Norte	247	13,7	13,7	36,5
	Occidente	581	32,2	32,2	68,7
	Oriente	204	11,3	11,3	80,0
	Sur	360	20,0	20,0	100,0
	Total	1804	100,0	100,0	



**RANGO DE EDAD:**

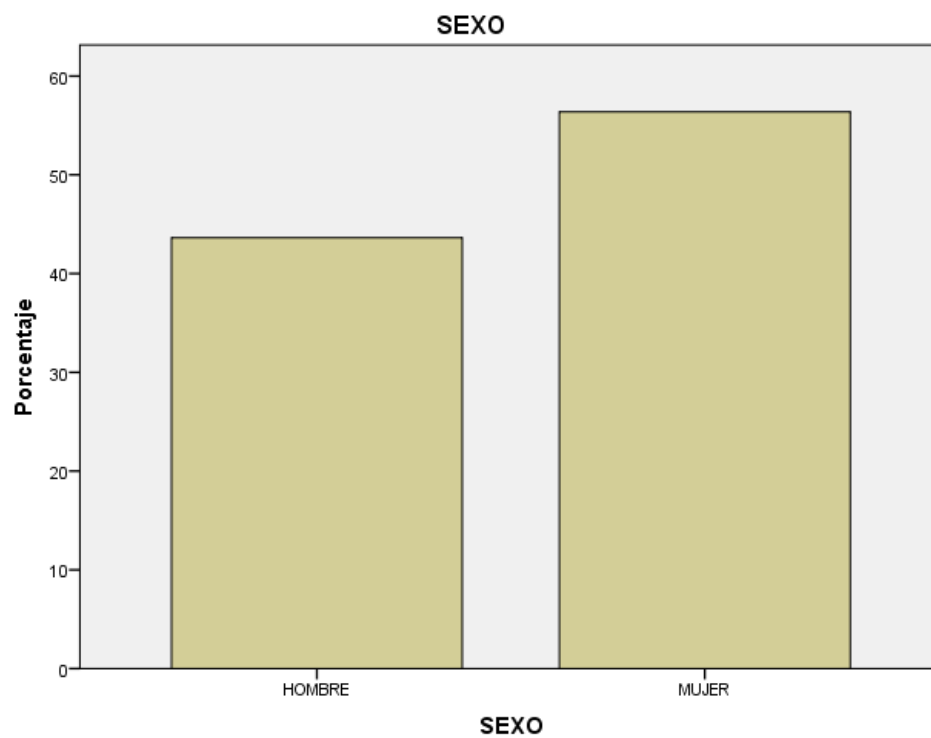
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos ADULTO (27 - 59 AÑOS)	1582	87,7	87,7	87,7
JOVEN (14 - 26 AÑOS)	73	4,0	4,0	91,7
MAYOR (MAYOR DE 60 AÑOS)	149	8,3	8,3	100,0
Total	1804	100,0	100,0	



El 87,7% de los beneficiados son adultos con edades entre los 27 y 59 años.

**SEXO:**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	HOMBRE	787	43,6	43,6	43,6
	MUJER	1017	56,4	56,4	100,0
	Total	1804	100,0	100,0	

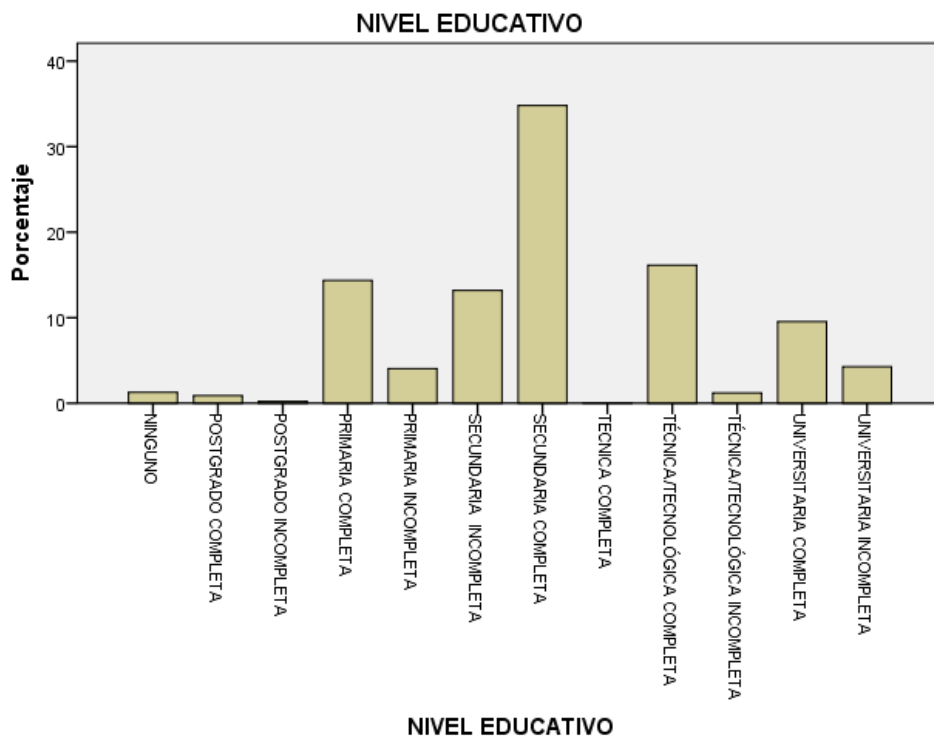


La proporción de hombres beneficiados es del 43,6% frente a un 56,4% de mujeres beneficiadas

**NIVEL EDUCATIVO**

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NINGUNO	23	1,3	1,3	1,3
POSTGRADO COMPLETA	16	,9	,9	2,2
POSTGRADO INCOMPLETA	4	,2	,2	2,4
PRIMARIA COMPLETA	259	14,4	14,4	16,7
PRIMARIA INCOMPLETA	73	4,0	4,0	20,8
SECUNDARIA INCOMPLETA	238	13,2	13,2	34,0
Válidos SECUNDARIA COMPLETA	628	34,8	34,8	68,8
TECNICA COMPLETA	1	,1	,1	68,8
TÉCNICA/TECNOLÓG ICA COMPLETA	291	16,1	16,1	85,0
TÉCNICA/TECNOLÓG ICA INCOMPLETA	22	1,2	1,2	86,2
UNIVERSITARIA COMPLETA	172	9,5	9,5	95,7
UNIVERSITARIA INCOMPLETA	77	4,3	4,3	100,0
Total	1804	100,0	100,0	

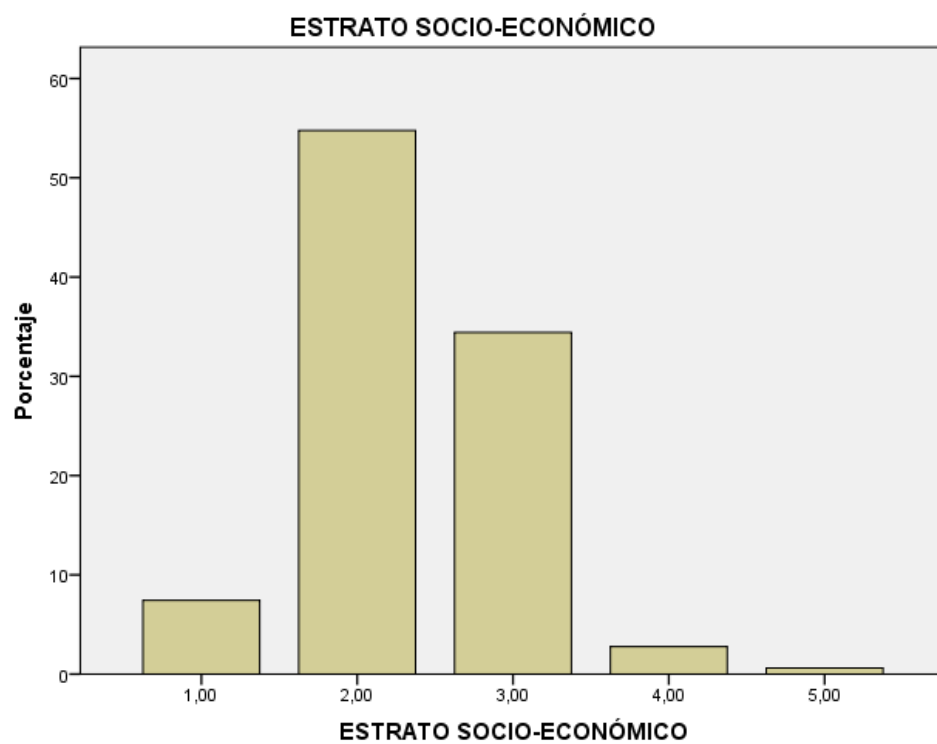




Demasiadas clases es necesario reagrupar

#### ESTRATO SOCIOECONÓMICO:

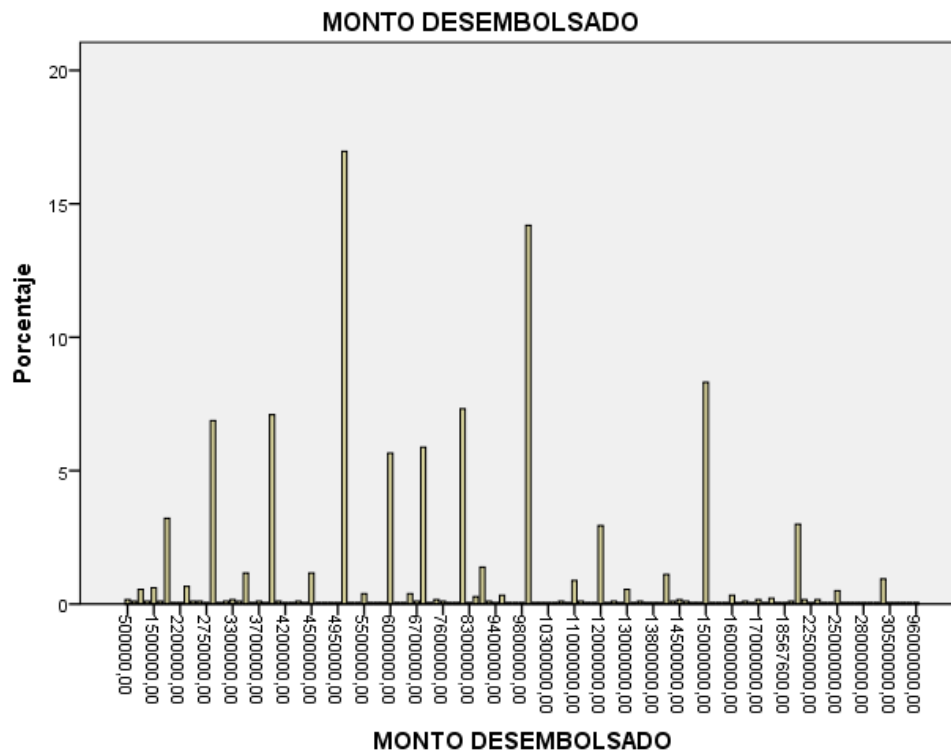
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	7,4	7,4	7,4
	2,00	54,8	54,8	62,2
	3,00	621	34,4	96,6
	4,00	50	2,8	99,4
	5,00	11	,6	100,0
	Total	1804	100,0	



**MONTO DESEMBOLSADO:**

N	Válidos	1804
	Perdidos	0
Media		8412530,4490
Mediana		7000000,0000
Moda		5000000,00
Desv. típ.		5968508,98318
Varianza		35623099482259,210
Asimetría		3,599
Error típ. de asimetría		,058
Curtosis		34,440
Error típ. de curtosis		,115
Rango		95500000,00
Mínimo		500000,00
Suma		15176204930,00

Coefficiente de variación
Media centrada
70,9%

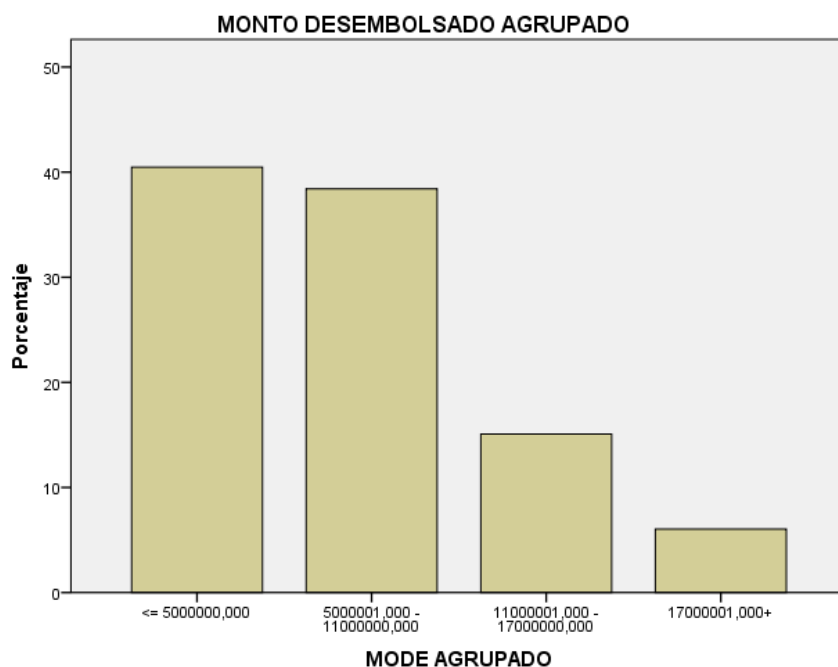


El Coeficiente de variación centrado en la media es de 70,9% lo que nos indica que existe alta variabilidad y por tanto la media no es representativa de la población, se plante categorizar la variable de la siguiente forma:

#### MONTO DESEMBOLSADO AGRUPADO

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<= 5000000,000	730	40,5	40,5	40,5
5000001,000 - 11000000,000	693	38,4	38,4	78,9
Válidos 11000001,000 - 17000000,000	272	15,1	15,1	94,0
17000001,000+	109	6,0	6,0	100,0
Total	1804	100,0	100,0	

Se agrupan los desembolsos en cuatro categorías



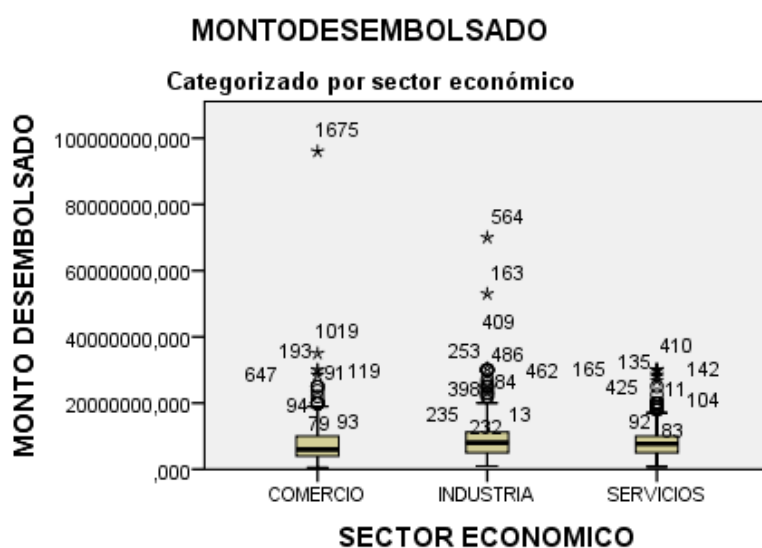
## APÉNDICE B

### RELACIONES DE DEPENDENCIA

#### MONTO DESEMBOLSADO

Se realiza un análisis entre la variable MONTO DESEMBOLSADO, por ser la variable de interés, y algunas de las variables explicativas mediante diagramas de caja, gráficas de barra y el estadístico chi-cuadrado de Pearson para determinar si existen relaciones de dependencia.

#### MONTO DESEMBOLSADO CATEGORIZADO POR SECTOR ECONÓMICO



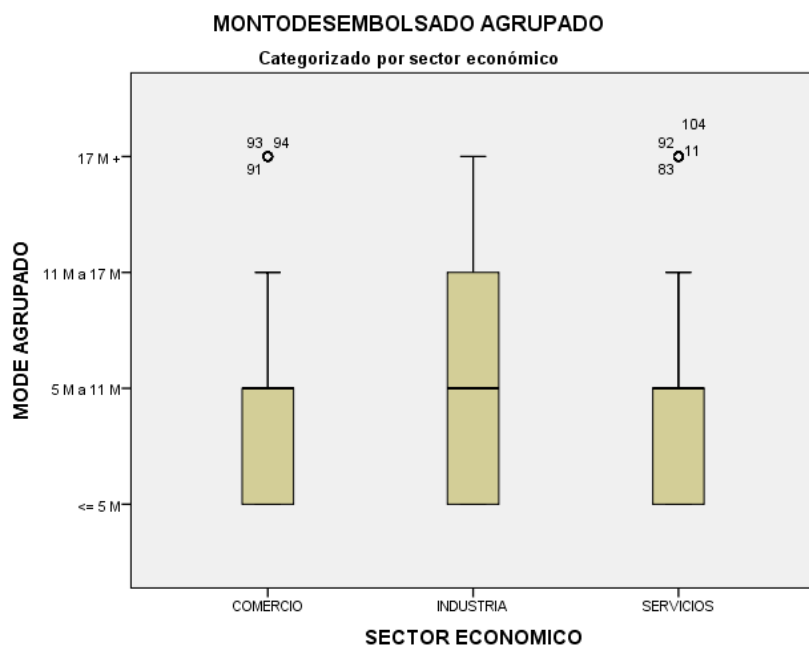


Tabla de contingencia MODE AGRUPADO \* SECTOR ECONÓMICO  
Recuento

		SECTOR ECONÓMICO			Total
		COMERCI O	INDUSTRI A	SERVICIO S	
MODE AGRUPADO	<= 5000000,000	315	266	149	730
	5000001,000 - 11000000,000	225	300	168	693
	11000001,000 - 17000000,000	81	134	57	272
	17000001,000+	23	55	31	109
	Total	644	755	405	1804

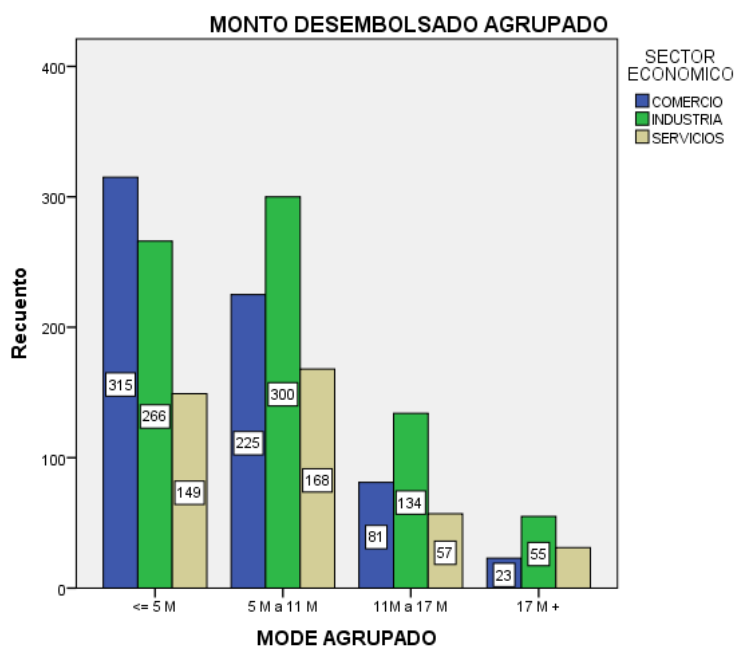
#### Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	37,870 <sup>a</sup>	6	,000
Razón de verosimilitudes	38,313	6	,000
N de casos válidos	1804		

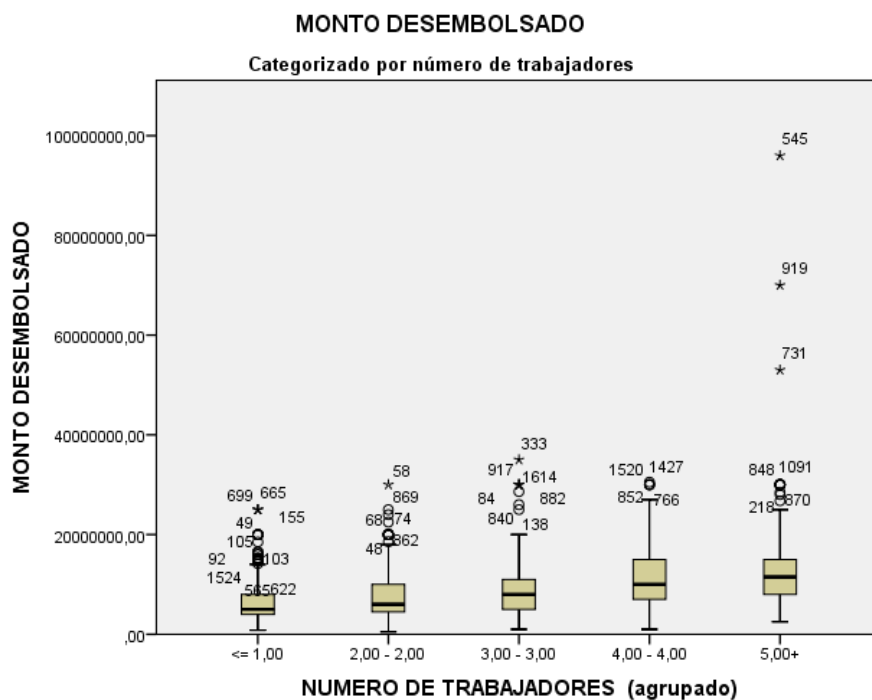
a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 24,47.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables

SEEC y MODE están relacionadas



#### 1.1.1.1 MONTO DESEMBOLSADO CATEGORIZADO POR NÚMERO DE TRABAJADORES



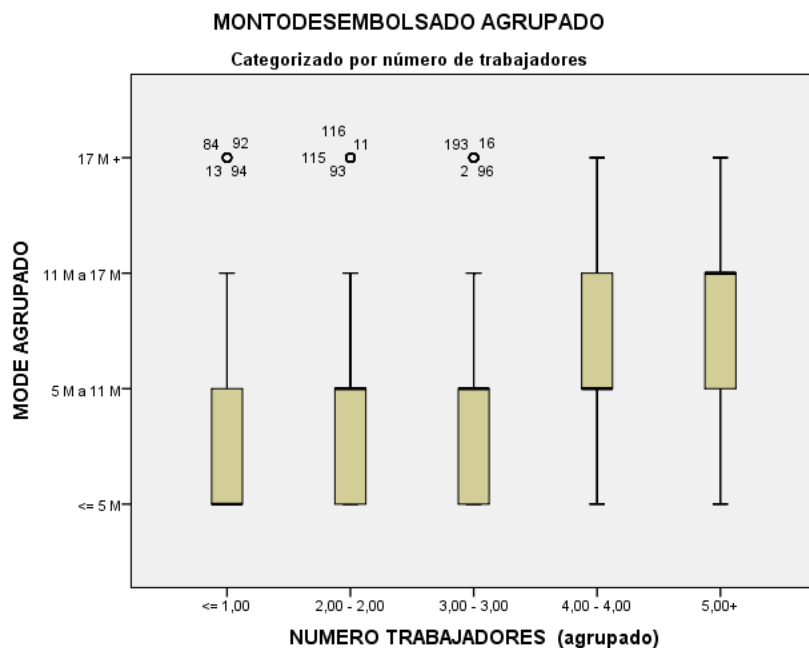


Tabla de contingencia MODE AGRUPADO \* NUMERO TRABAJADORES (agrupado)

		NUMERO TRABAJADORES (agrupado)					Total
		<= 1,00	2,00 - 2,00	3,00 - 3,00	4,00 - 4,00	5,00+	
MODE AGRUPADO	<= 5 M	333	244	90	36	27	730
	5 M a 11 M	211	223	114	70	75	693
	11 M a 17 M	54	66	46	45	61	272
	17 M +	10	17	19	22	41	109
Total		608	550	269	173	204	1804

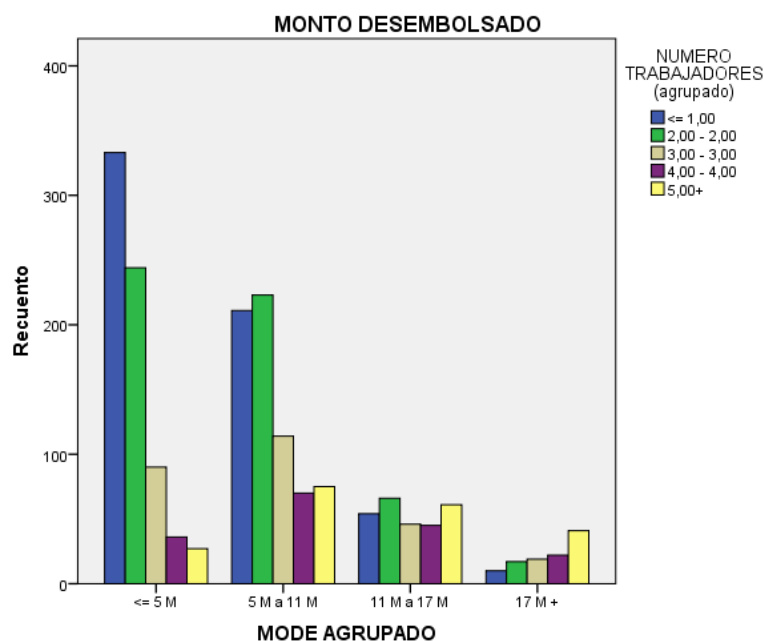
#### Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	264,664 <sup>a</sup>	12	,000
Razón de verosimilitudes	253,610	12	,000
Asociación lineal por lineal	250,255	1	,000
N de casos válidos	1804		

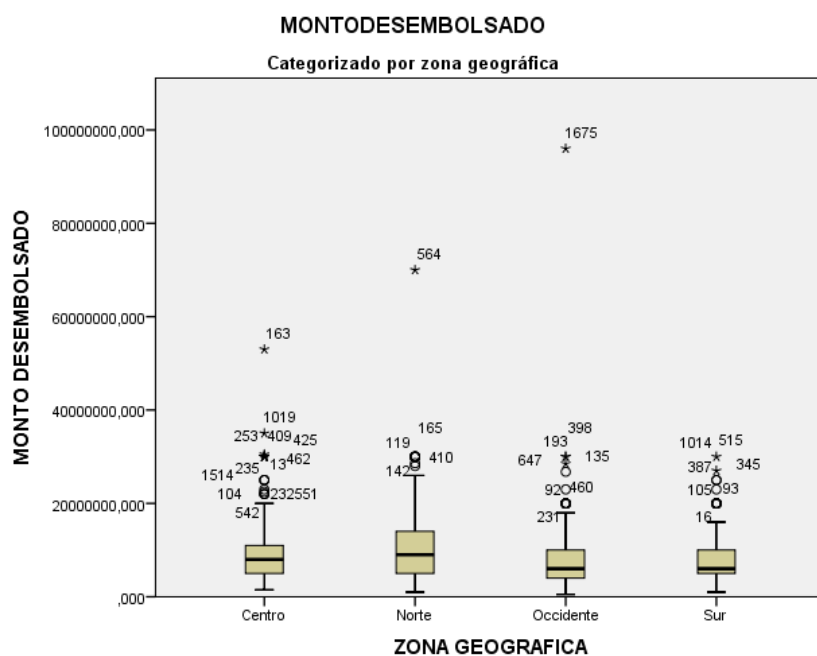
a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 10,45.



Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables NUTR y MODE están relacionadas



#### 1.1.1.2 MONTO DESEMBOLSADO CATEGORIZADO POR ZONA GEOGRÁFICA



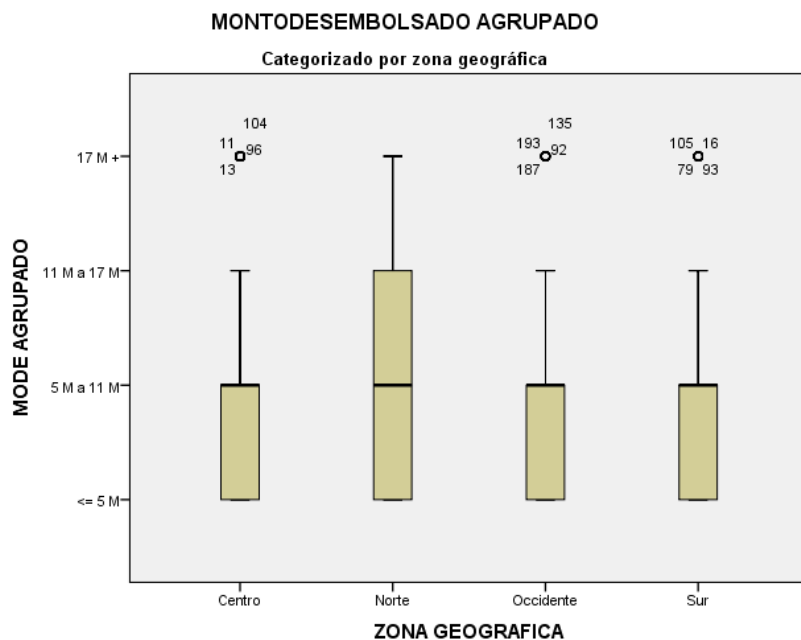


Tabla de contingencia MODE AGRUPADO \* ZONA GEOGRÁFICA

		ZONA GEOGRÁFICA				Total
		Centro	Norte	Occidente	Sur	
MODE AGRUPADO	<= 5 M	161	132	201	236	730
	5 M a 11 M	164	167	146	216	693
	11 M a 17 M	70	94	56	52	272
	17 M +	35	40	18	16	109
Total		430	433	421	520	1804

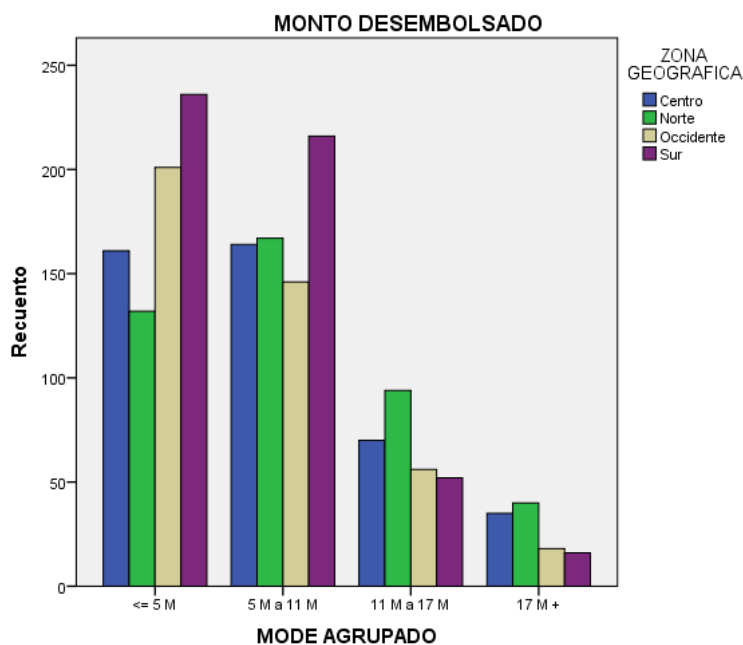
#### Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	66,116 <sup>a</sup>	9	,000

Razón de verosimilitudes	67,036	9	,000
N de casos válidos	1804		

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 25,44.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables ZONA GEOGRÁFICA y MODE están relacionadas



#### 1.1.1.3 MONTO DESEMBOLSADO CATEGORIZADO POR RANGO DE EDAD

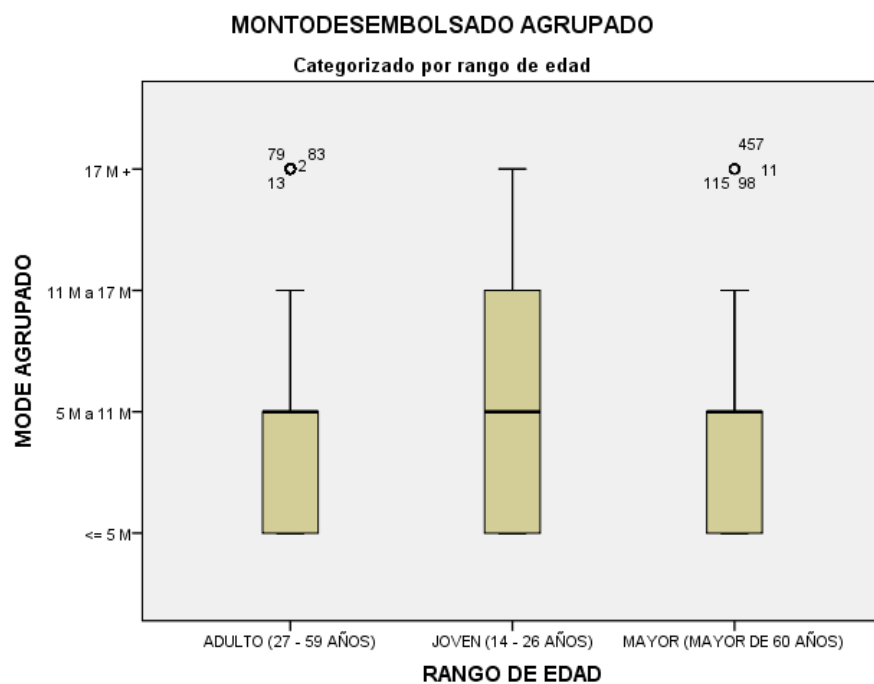
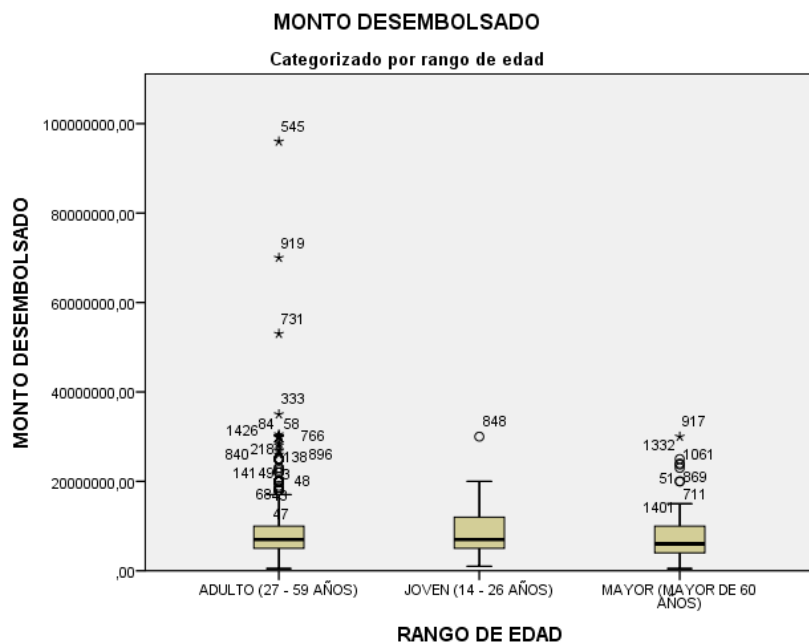


Tabla de contingencia MODE AGRUPADO \* RANGO DE EDAD

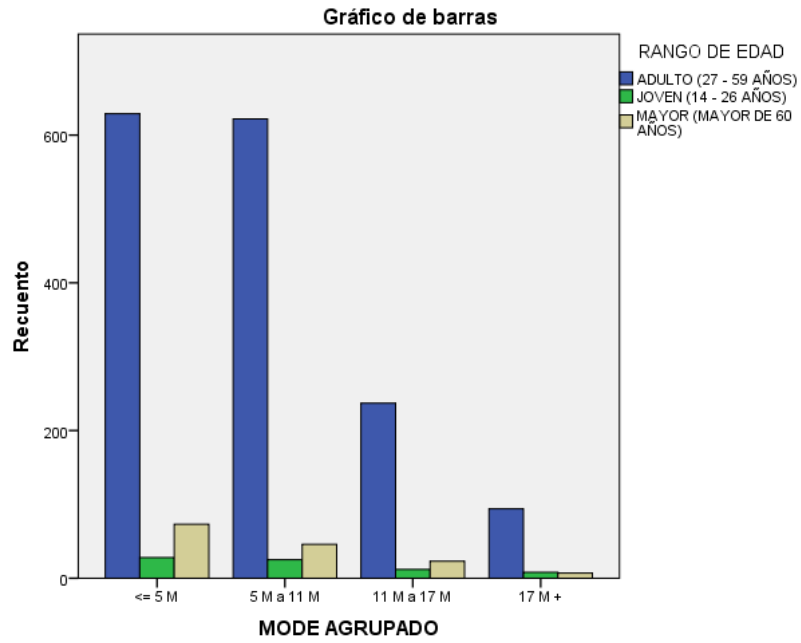
		RANGO DE EDAD			Total
		ADULTO (27 - 59 AÑOS)	JOVEN (14 - 26 AÑOS)	MAYOR (MAYOR DE 60 AÑOS)	
MODE AGRUPADO	<= 5 M	629	28	73	730
	5 M a 11 M	622	25	46	693
	11 M a 17 M	237	12	23	272
	17 M +	94	8	7	109
Total		1582	73	149	1804

### Pruebas de chi-cuadrado

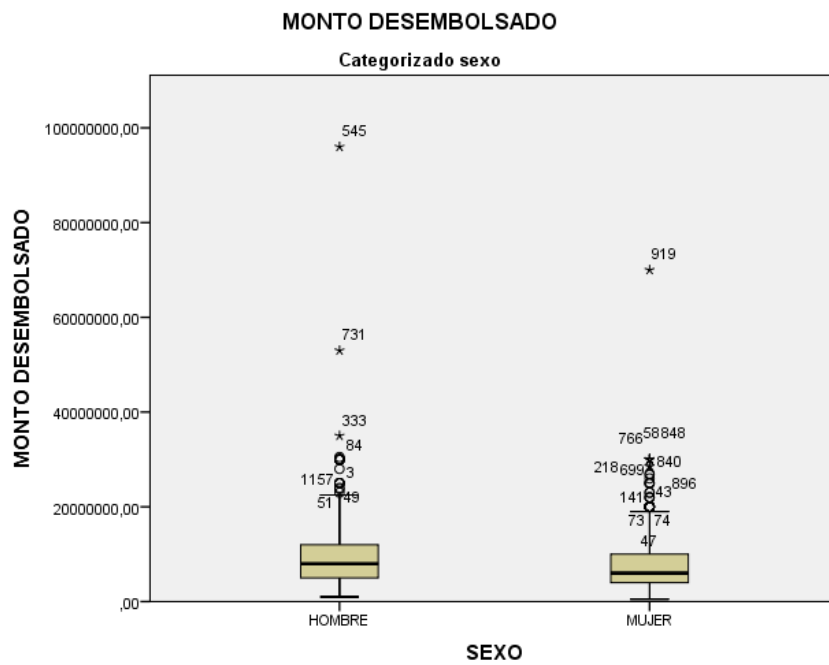
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	9,330 <sup>a</sup>	6	,156
Razón de verosimilitudes	8,794	6	,185
N de casos válidos	1804		

a. 1 casillas (8,3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 4,41.

Dado que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy alta, se decide validar la hipótesis de independencia y concluir que las variables EDAD y MODE no están relacionadas



#### 1.1.1.4 MONTO DESEMBOLSADO CATEGORIZADO POR SEXO



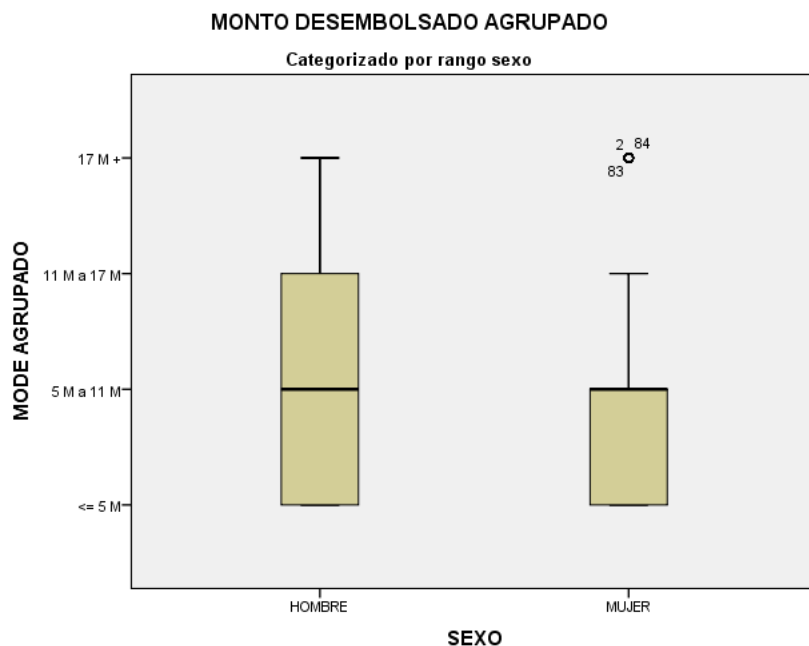


Tabla de contingencia MODE AGRUPADO \* SEXO

		SEXO		Total
		HOMBRE	MUJER	
MODE AGRUPADO	<= 5 M	236	494	730
	5 M a 11 M	326	367	693
	11 M a 17 M	160	112	272
	17 M +	65	44	109
Total		787	1017	1804

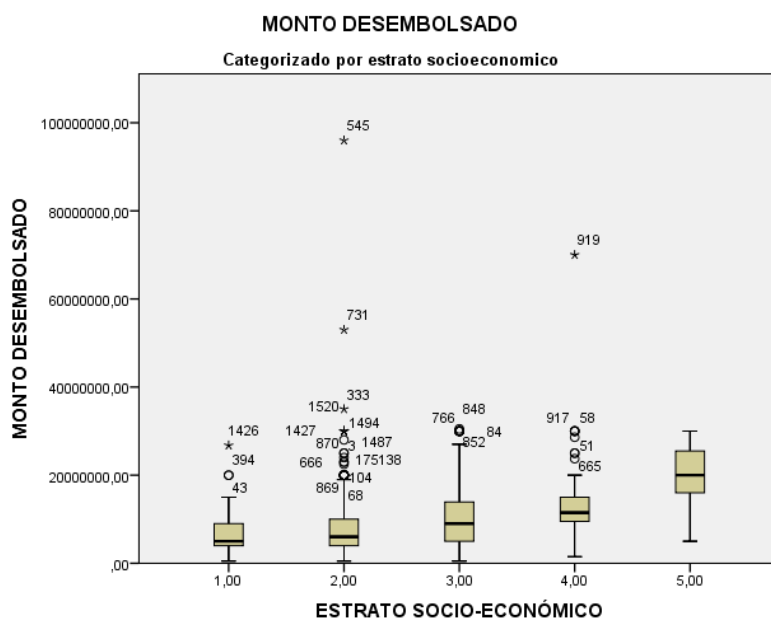
Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	78,071 <sup>a</sup>	3	,000
Razón de verosimilitudes	78,793	3	,000
N de casos válidos	1804		

- a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 47,55.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables SEXO y MODE están relacionadas.

#### 1.1.1.5 MONTO DESEMBOLSADO CATEGORIZADO POR ESTRATO SOCIOECONÓMICO





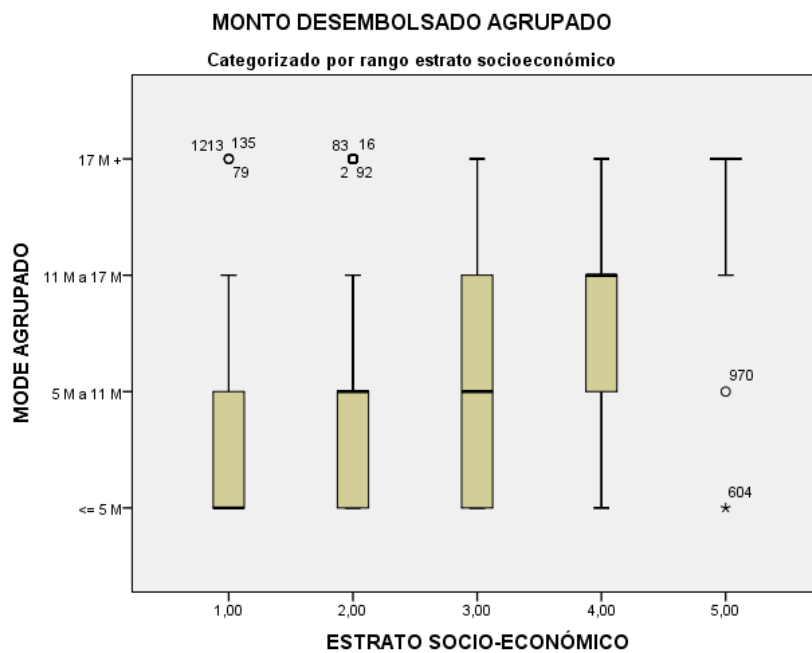
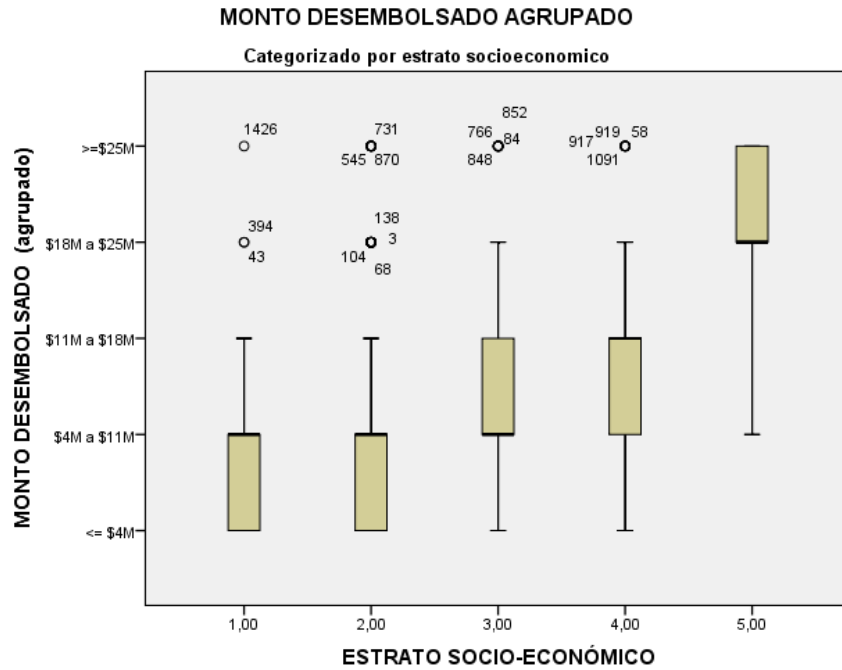


Tabla de contingencia MODE AGRUPADO \* ESTRATO SOCIO-ECONÓMICO

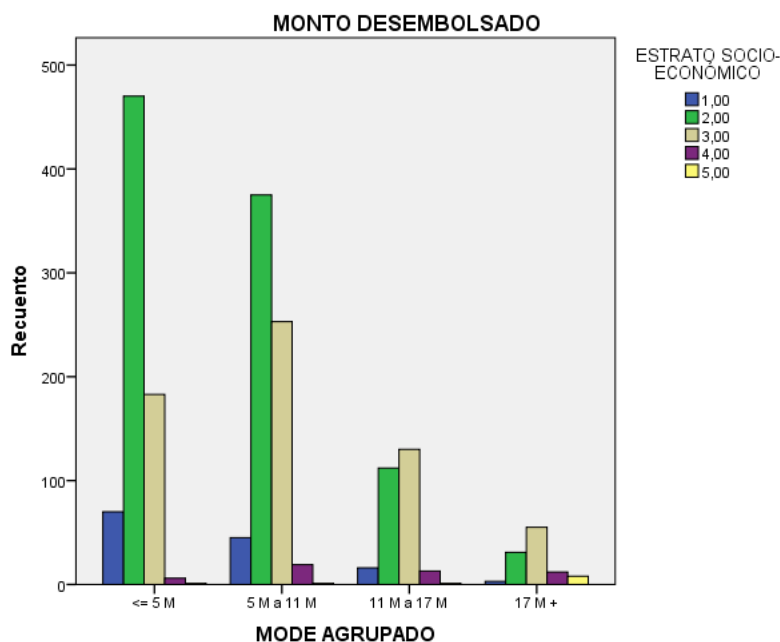
		ESTRATO SOCIO-ECONÓMICO					Total
		1,00	2,00	3,00	4,00	5,00	
MODE AGRUPADO	<= 5 M	70	470	183	6	1	730
	5 M a 11 M	45	375	253	19	1	693
	11 M a 17 M	16	112	130	13	1	272
	17 M +	3	31	55	12	8	109
Total		134	988	621	50	11	1804

## Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	213,552 <sup>a</sup>	12	,000
Razón de verosimilitudes	155,020	12	,000
Asociación lineal por lineal	134,946	1	,000
N de casos válidos	1804		

a. 5 casillas (25,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es ,66.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables ESEC y MODE están relacionadas.



## 1.1.2 VARIABLES EXPLICATIVAS

### 1.1.2.1 SEXO VS ESTRATO SOCIOECONÓMICO

Tabla de contingencia

		ESTRATO SOCIO-ECONÓMICO					Total
		1,00	2,00	3,00	4,00	5,00	
SEXO	HOMBRE	54	404	300	24	5	787
	MUJER	80	584	321	26	6	1017
Total		134	988	621	50	11	1804

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	9,551 <sup>a</sup>	4	,049
Razón de verosimilitudes	9,532	4	,049
N de casos válidos	1804		

a. 1 casillas (10,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 4,80.

Se acepta la hipótesis de independencia por tanto las variables SEXO y ESEC no están relacionadas

## 1.1.2.2 SEXO VS SECTOR ECONÓMICO

Tabla de contingencia SEXO \* SECTOR ECONÓMICO

		SECTOR ECONÓMICO				Total
		AGROPECUARIO	COMERCIO	INDUSTRIA	SERVICIOS	
SEXO	HOMBRE	5	283	330	169	787
	MUJER	9	361	411	236	1017
Total		14	644	741	405	1804

## Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	1,224 <sup>a</sup>	3	,747
Razón de verosimilitudes	1,233	3	,745
N de casos válidos	1804		

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 6,11.

Se acepta la hipótesis de independencia por tanto las variables SEXO y SECTOR ECONÓMICO no están relacionadas

## 1.1.2.3 SEXO VS NÚMERO DE TRABAJADORES

Tabla de contingencia SEXO \* NUMERO TRABAJADORES (agrupado)

		NUMERO TRABAJADORES (agrupado)					Total
		<= 1,00	2,00 - 2,00	3,00 - 3,00	4,00 - 4,00	5,00+	
SEXO	HOMBRE	214	233	127	84	129	787
	MUJER	394	317	142	89	75	1017
Total		608	550	269	173	204	1804

## Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	52,930 <sup>a</sup>	4	,000
Razón de verosimilitudes	53,020	4	,000
N de casos válidos	1804		

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 75,47.

Puesto que la probabilidad (denominada nivel crítico o nivel de significación observado) es muy pequeña, se decide rechazar la hipótesis de independencia y concluir que las variables NUTR y SEXO están relacionadas.

